UNIVERSITY OF BIRMINGHAM

HAZARD UNCERTAINTY AND RELATED DAMAGE POTENTIALS OF EXTRA–TROPICAL STORMS

by

MICHAEL ALEXANDER WALZ

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Geography, Earth and Environmental Sciences

College of Life Sciences

University of Birmingham

November 2018

Abstract

Extra–tropical winter windstorms are among the most loss–intensive natural hazards in Europe. This thesis is dedicated to advance the understanding of these hazardous events and their uncertainty in various aspects. These aspects include the serial clustering and spatial variability of storm events, the seasonal predictability of extreme wind speeds associated with windstorms and an impact assessment of windstorms both in a climatological as well as from a loss–related perspective. The recurring element in all studies are large–scale drivers (e.g. North Atlantic Oscillation - NAO) which are linked to different features of extra–tropical windstorms, e.g. the inter–annual variability. It can be shown that large–scale drivers are able to explain a considerable amount of variability of windstorms. Seasonal forecast ensemble hindcasts are used to create a physical consistent virtual reality of more than 1500 years. Thus, the uncertainty of these extreme events can be estimated more accurately compared to using century–long reanalysis. This large sample size can also be used to estimate potential extremes with respect to intensity and severity of windstorms more accurately. The findings of these studies are presented in five scientific papers which are included as five chapters in this submitted thesis.

Acknowledgements

Firstly I would like to thank my supervisor Gregor Leckebusch for his support and academic guidance throughout the three years of my PhD. I really enjoyed our "coffee mornings" which often turned into fruitful discussions on my research. Not uncommonly I came out of our meetings with new and fresh ideas on what to tackle next. Due to his encouragement I also had the chance to visit numerous conferences and workshops all around the world which I personally found extremely useful. My travels included a stay at the Climate Change Research Centre at the University of New South Wales in Sydney. Here I would particularly like to thank Markus Donat for hosting me there and also for our engaging discussions, particularly on atmospheric predictability. Further I would like to thank all my other co–authors, Daniel Befort, Nicolas Kircher–Bossi, Tim Kruschke, Henning Rust and Uwe Ulbrich, which helped me to produce five scientific papers within my three years as a PhD student. I would also like to thank the School of Geography, Earth and Environmental Sciences (GEES) at the University of Birmingham for giving me the chance to pursue this project in the first place. Here at GEES I was lucky to meet many great people which certainly made my stay here in Birmingham much more fun. Thanks to the "Berlin enclave" aka Simon and Daniel, Colin, Sally, Vasilis, Clemens and also all the members of the glorious Bayern Abrantes team which unfortunately did not survive the loss of our star players. I would also like to thank all my friends back in Germany. Many of them came to visit me over the last three years. It was always great to have them over and take a short break from university life. Last but not least I would also like to thank my parents who have always supported me in my decisions and thereby helped me to achieve my goals.

Contents

List of Figures

List of Tables

Acronyms

AIC	An (Akaike) information criterion
AMO	Atlantic Meridional Oscillation
AS	Active season (clustering)
BI	British Isles
CE	Central Europe
DI-SSI	Distribution-independent Storm Severity Index
DJF	December–January–February
EA	East Atlantic Pattern
EAWR	East Atlantic/Western Russia Pattern
ED	Explained Deviance
EM	Expectation–Maximization Algorithm
ECMWF	European Centre of Medium Range Forecasts
ENSO	El Niño-Southern Oscillation
EOF	Empirical Orthogonal Function
EPNP	East Pacific–North Pacific Pattern
EPS	Ensemble Prediction System
ETC	Extra–tropical Cyclone
EUR	Euro (currency)
GEBE	Germany and Benelux

GLM	Generalised Linear Model
GPD	Generalised Pareto Distribution
hPa	Hektopascal
HR	Hit rate (clustering)
IAS	Inactive Season (clustering)
ISMR	Indian Summer Monsoon region
MSLP	Mean sea level pressure
NAO	North Atlantic Oscillation
NH	Northern Hemisphere
PNA	Pacific–North American Pattern
PI	Predictive Information
POL	Polar Pattern
POT	Peak Over Threshold
PP	Predictive Power
PSC	Primary Storm Cluster
PV	Potential Vorticity
QBO	Quasi–biennial Oscillation
RMSE	Root Mean Square Error
SC	Scandinavia
SCA	Scandinavian Pattern
SSI	Storm Severity Index
SST	Sea Surface Temperature
TC	Tropical Cyclone
TNH	Tropical Northern Hemisphere Pattern
USD	US Dollar
VIF	Variance Inflation Factor
WP	West Pacific Pattern

Chapter 1

Preface

"If you want to see the sunshine, you have to weather the storm."

– Frank Lane, American Baseball Executive

On November 16th 1854, during the Crimean War[1], the British Captain William Powell Richards wrote a letter to his aunt in which he reported:

"Since I commenced this we have had a terrible storm of wind, rain, and snow, giving us a taste of what the winter will be. We have suffered great loss, in the first place, eight store ships, three of them steamers, were wrecked outside Balaklava, and 300 lives lost, not only this, they contained almost all the winter clothing sent for us, so God knows what we shall do, as it will take at least two months to get more from England. It is calculated that property to the amount of three millions was lost, amongst which I will mention 9,000 gallons of rum and from four to six million rounds of mine and musket armaments, immense quantities of beef, pork, biscuit, hay, barley, sugar, and clothes for the troops, a large quantity of siege ammunition, which was much wanted, in addition to this the wind blew a perfect hurricane, levelled nearly every tent, mine amongst the number, the consequence was that everything got wet in our tents, and half our things spoilt, it was bitterly cold also."

As reported by Captain Richards the storm of November 1854 represented a major setback during the siege for Sevastopol for the British and French Army, in particular because of the loss of 30 vessels. Among the ones lost were the British *HMS Black Prince* as well as the French battle ship *Henri IV*. As a direct consequence to this tragic loss the famous astronomer Urbain Le Verrier was assigned to study the development of this storm by the French government. The main conclusion of his report was that a (telegraphic) network of meteorological observations could

[1] As Jonathan Walford notes this war was apparently also the hour of birth of a knitted garment named "Cardigan" named after Thomas Brudenell, the 7th Earl of Cardigan "who distinguished himself while wearing a cardigan style (front button, waist length) wool jacket when he lead the Charge of the Light Brigade at Balaclava 1854" (Walford, 2013)

be used to monitor severe storms and thereby issue warnings of in the future. By 1856 Le Verrier had built a network of observations from 19 weather stations (15 in France and 4 abroad) which were linked by the electric telegraph (Encyclopedia, 2008). This network was used operationally to issue warnings to seafarers without any attempt to actually predict the weather, however. This network of observations represented one of the earliest attempts of storm tracking and initiated the first governmental meteorological forecasting service (Walker, 2011; Lindgrén and Neumann, 1980; Hibbert, 1961; Landsberg, 1954).

1.1 Outline of the thesis

The following sections (Chapter 1.2-1.5) provide an introduction to the topics covered by this thesis and try to connect different aspects regarding European winter windstorms. The main body of this thesis (Chapter 2-6) consists of three published papers (Walz et al., 2017, Walz et al., 2018a and Walz et al., 2018b), and two papers currently (March 9, 2019) under review (Walz et al., 2018c/2019 and Walz et al., 2019). The chapters consist of LaTeX edited versions of the accepted/submitted manuscript of the respective article. Chapter 7 summarises the advances in the field of extra–tropical storms made by this thesis and suggests areas for future research. The thesis is concluded by the References (that include *all* references from *all* chapters) and the Appendix.

1.2 Severity quantification of windstorms

Even today, more than 160 years after the storm of 1854, there is still a need for research regarding the quantification of the severity or the impact of a windstorm. Captain Richards lists 300 casualties and property loss of around three million which is a surprisingly exact estimation of the impact that the storm had. As weather reporting, however, was in its infancy at that time Captain Richards had no way of estimating the meteorological "extremeness" or the hazard which the storm represented. The only thing that was available during that time was the Beaufort scale which was devised by the Irish hydrographer Francis Beaufort in 1805 (Beaufort, 1805). In 1838 the observation and classification of wind speeds in the Beaufort scale became a mandatory log entry for all ships of the Royal Navy (Wheeler, 2001). As the classification of winds into the Beaufort scale is purely based on observations, however, it is prone to subjectivity and also local effects. The impact of a Beaufort number 9 for example is described as "Slight structural damage (chimney pots and slates removed)" (WMO, 1970). On the meteorological side this is equivalent to a wind speed between 20.8 and 24.4 m/s. The impact of the highest Beaufort number (12) is simply described as "Devastation" (WMO, 1970) and is used for any wind higher than 32.7 m/s. The associated impact of these wind speeds however is purely based on experience and knowledge which could easily be 50 years old.

The question of how to conveniently quantify the severity of windstorms is still ongoing research. A first attempt was made by Lamb and Frydendahl (1991) who introduced the cube of a wind speed as the kinetic energy advected by the wind itself. Their approach was applied to (interpolated) station data and weather maps. Naturally this implicates uncertainties depending on the synoptic situation and in-

terpretation of the weather maps. Thus, the estimated area which they considered affected suffered from uncertainty as well. Leckebusch et al. (2008a) introduced a Storm Severity Index (SSI) that combines the meteorological hazard with the vulnerability of local conditions by scaling the wind speeds with the local 98^{th} percentile (V98) in a more objective way. This approach was initially introduced by Klawa and Ulbrich (2003) who found the best fit for their loss parametrisation using the highest 2% of the local wind speeds. The V98 can be seen as a proxy for the local resilience towards wind speeds. Thus this approach has been widely used in the loss modelling community. Schwierz et al. (2010) for example identified days of extreme gust winds by selecting days with wind speeds that were 30% larger than V98 ($V/V98 > 1.3$). Other studies have related the loss to the maximum wind which is however scaled by a local constant, thus also taking into account a local resilience (e.g. Prahl et al., 2012).

These previous indices use the tail of the wind speed distribution to quantify the severity of windstorms without assessing the shape of the tail of the distribution. Thus, depending on if wind speeds are distributed in a light– or heavy–tailed distribution, the resulting quantification is influenced by the amount of exceedances of the respective percentile. In particular this could mean that the severity of an event in areas with only a few exceedances could be overestimated compared to a region with frequent exceedances. To overcome this situation the first chapter of this thesis is dedicated to develop a severity quantification of European winter windstorms which integrates the shape of the upper tail of the distribution, thus the extreme wind speeds. This quantification can be regarded as a more climatological extremeness compared to the more impact related quantification of the SSI.

1.3 Variability of windstorms

As a result of the disastrous events for the French Army during the storm of November 1854 Urbain Le Verrier introduced one of the first "now-casting" networks over a considerable spatial extent using telegraph lines. This was around the same time when the British Vice-Admiral Robert FitzRoy was appointed "Meteorological Statist to the Board of Trade" which would later become the UK Met Office. As one of his first official acts he asked captains of ships to report meteorological data such as temperature or pressure. Initially these measurements were only used for data collection, however with more telegraphs available the readings could also be transmitted almost in real time. He also introduced barometers on many big harbours around the UK (Mellersh, 1968). The main purpose of these networks was to issue early warnings to entities at risk (i.e. vessels). The implementation of such a facility shows that government officials and scientists likewise had understood that events as the 1854 storm were not unique phenomenons but rather periodic hazards that would affect Europe afresh every year. The fact of the rather large spatial extent of the observation network also showed that people of their time were aware that winterstorms are a) affecting areas across several hundred kilometres and b) feature a spatial variability, thus they tend to affect different regions every year. Up until present day research, the question of the variability of winter windstorms on a temporal as well as on a spatial scale remains partly unanswered.

There have been numerous studies investigating the inter-annual variability of winter windstorms for different European regions on time scales up to 120 years. Cusack (2013) for example examined the storminess in the Netherlands over the last 100 years concluding that losses from windstorms have remained stable over the period,

however they do show a cycle of around 50 years. A study based on mean sea level pressure observations in Sweden by Bärring and von Storch (2004) showed that the 1860's and 70's and the 1980's and 90's represented stormy decades in particular. They do note however that both of these periods are still close to the long-term average, thus not exceptionally high for Sweden. A similar study for Switzerland was published by Welker and Martius (2014) who blended observations with the NOAA 20CR (Compo et al., 2011) reanalysis. They found a periodicity of extreme winds associated with windstorms between 36-47 years. Moving from observations towards reanalysis, Befort et al. (2016) compared the inter-annual variability of windstorms in the century-long reanalysis data sets ERA-20C (Poli et al., 2016) with NOAA 20CR for the 20^{th} century. They could identify a trend in the observed windstorms which they regard as an artefact of the data set however as it is different to all other reanalysis widely used for windstorm analysis. A special type of inter-annual variability is represented by the serial clustering of windstorms (Mailier et al., 2006, Vitolo et al., 2009 or Pinto et al., 2013). This phenomenon is particularly interesting for the impact community hence the actuarial industry.

The question that naturally arises when investigating the variability is the one concerning the drivers of this variability. Thus, is there an atmospheric/oceanic mechanism that can explain the variability of windstorms on time scales from months to decades? Or in a more technical sense: Is there a driver to be identified which a) can help to understand the variability of the storms and b) can be used as a predictor (in a statistical sense) for the forecast of windstorms on various time scales? In order to link these three components of variability, clustering and potential predictors chapter three is designed to improve the understanding of windstorm clustering as

well as to identify geophysical drivers that explain this observed variability of windstorms over the last 100 years. Chapter five of this thesis addresses the question of spatial variability, thus which areas are affected more/less by extreme windstorms and it also tries to give an insight on the "worst-case scenario" in terms of extreme windstorms for different European regions.

The variability (both spatial or temporal) is due to mechanisms on different scales: Starting from a small (synoptic) scale including potential vorticity (PV) anomalies (Hoskins, 1974; Hoskins et al., 1985) and the theory of baroclinicity (Charney, 1947) to a continental scale including the NAO (Hurrell, 1995) or other large–scale patterns up to the global scale including the jet stream (e.g. Branstator, 2002). Understanding the interplay between those mechanisms is key to understand the variability of extra–tropical storms. Overall this thesis is focused more on the larger scales (continental to global) as the idea is to gain an understanding of windstorms as hazards and their associated impacts and uncertainties. However there is a lot of literature on the actual "physical" process of the intensification of extra–tropical cyclones e.g. Simmons and Hoskins (1979), Shapiro et al. (1999), Pinto et al. (2007) or Tierney et al. (2018). Rivière et al. (2012) for example could show that for windstorm Xynthia the PV gradient anomaly was responsible for a poleward shift of the cyclone track.

1.4 Predictability of extreme wind speeds associated with windstorms

Although Urbain Le Verrier never actually predicted weather, he came to the conclusion that warnings of the storm in November 1854 could have been issued at least a day in advance if observations from around the area had been available (Landsberg, 1954). Given that meteorology as a science was in its infancy and that data was virtually non-existent this can be regarded as quite a bold opinion at that time. Nowadays many decisions in everyone's daily routine are based on the weather forecast. Thus it almost seems inconceivable that people 160 years ago were completely left in the dark what the weather within the next hours, let alone the next days or weeks was going to be like. Certainly meteorological research has developed rapidly over the last century and there have been numerous studies on the forecast of both tropical (Kurihara et al., 1995; Rappaport et al., 2009; Tsai et al., 2011; Zhang et al., 2017) and extra-tropical cyclones (Froude et al., 2007; Froude, 2010). Predictions in these studies are made for a time scale of about 7–10 days. Over the last decade there has been a rising demand of "long-term" forecasts covering time-scales from months over years up to decades. Due to the chaotic nature of the atmosphere it is however impossible to skilfully predict the weather for more than 7 days in a deterministic sense. This chaotic nature which is also referred to as "Butterfly effect" goes back to Lorenz (1963) who paved the way for modern weather forecast with his study.

As deterministic forecasting is not possible for those longer time scales, science has moved to so called probabilistic forecasts realised by ensemble predictions. Although deterministic forecasts are often probabilitic as well, the main difference is that for

deterministic forecasts, the probability distributions are for states at a given time whereas for long–term predictions the probability distributions are used to generate statistics (mean, exeedance probabilities) over an extended period. For example instead of saying the average temperature next March will be 5.8° C, a probabilistic forecast would say there is, e.g., a 35% chance that next March will be warmer than normal (the median over a given period).

To meet the demand for predictions that exceed the time scales of the conventional weather forecast, a large seasonal prediction community has formed over the last decade. One of the first initiatives was set up by Palmer et al. (2004). Since then there have been numerous studies on the evaluation of seasonal ensemble prediction systems with regards to all sorts of meteorological variables: Sahai et al. (2015) for example found significant skill for the ECMWF System 4 (Molteni et al., 2011) for the precipitation associated with the Indian summer monsoon in the Indian Summer Monsoon Region (ISMR). Kim et al. (2012) also investigated System 4 intensively. They found a cold bias in the equatorial Pacific but a warm bias in the North Pacific and the North Atlantic. Generally they found that the skill for the 2m temperature and precipitation anomalies is greater in the tropics than in the extra-tropics. Additionally the skill for the 2m temperature and precipitation in the tropics appears to be larger in strong El Nino years.

The UK Met Office Global Seasonal forecast System 5 (GloSea 5; MacLachlan et al., 2015) showed significant skill for predicting the North Atlantic Oscillation (NAO; e.g. Hurrell, 1995) for the European winter period (Scaife et al., 2014). Della-Marta et al. (2010) investigated the skill of ECMWF System 3 (Anderson et al., 2007;

which is the predecessor of System 4) and found skill for the 95^{th} percentile of wind speeds for the first lead month but none thereafter, however. To the knowledge of the author of this thesis however, Della-Marta et al. (2010) is the only study dedicated to investigate the skill of the actual extreme wind speeds in seasonal forecast systems. Renggli et al. (2011), however were one of the first to investigate the skill of windstorm events. Also there has been no examination of the drivers of extreme wind speeds so far, especially for the "seasonal-forecast-internal" drivers. Chapter 4 tries to fill this gap by firstly investigating the predictive skill of a seasonal forecast system and secondly link the predicted extreme wind speeds with large scale atmospheric drivers. By doing so potential shortcomings of the seasonal forecast system might be discovered and addressed for future seasonal forecast products.

1.5 Impact of winter windstorms

In his letter in 1854 Captain Richards gave a surprisingly detailed estimation of how much damage the storm in 1854 had caused. Even more than 160 years after this letter was written, the question of how much damage (or loss) is caused by windstorms is still a very difficult one to answer. In the actuarial sector loss (from any natural peril) is usually made up of four different modules (Schwierz et al., 2010): Hazard, Exposure, Value Distribution and Cover conditions. Especially the second and third items on that list bear a large regional uncertainty. The resilience against a windstorm for example is higher in countries where they occur more frequent. Furthermore the distribution of the entities at risk (houses, cars, factories...) has to be known on a very small scale in order to give realistic loss assessments. The main problem when trying to calculate losses associated with windstorms (or other natural perils) in the academic world is the lack of available (insured) loss data to calibrate the damage function of a loss model.

There are a few studies using losses for the UK and especially Germany (Leckebusch et al., 2007, Held et al., 2013 or Donat et al., 2011). However, the reported losses cover only Germany and are on a time scale of around 10-15 years. Recently there have been a few efforts to make the world of loss modelling a bit more accessible for academia. The OASIS loss modelling framework (`http://www.oasislmf.org`) for example provides an open source platform to use a selection of operational loss models developed by various companies. Another initiative is the open source natural catastrophe (NatCat) loss model *climada* (Bresch and Mueller, 2017) which provides software to create probabilistic hazard sets, exposure data and to ultimately calculate losses based on damage functions. It is a very extensive piece of software

which also enables the user to investigate the impact of adaptation measures (e.g. change of building code, flood protection...).

Besides the exposure and the value distribution there is also a lot of uncertainty regarding the hazard of windstorms. Recent reanalysis efforts are on a scale of roughly one hundred years (e.g. ERA–20C; Poli et al., 2016). Given the scarcity of highly loss–intensive windstorms this goes along with a large uncertainty regarding the loss especially when going from the European to the country scale. In order to overcome the lack of storms a common approach is to create probabilistic hazard sets that use perturbations of actual events (e.g. Schwierz et al., 2010). Events created that way, however, do not necessarily take into account all the environmental constraints. Thus, some of the windstorms might physically not be possible.

Chapter 6 is designed to create a physically consistent windstorm hazard set from seasonal forecast data and to estimate annual losses with the help of the NatCat tool *climada*. Due to the setup of the seasonal forecast model more than 1500 model years are available. Additionally this chapter further investigates the question of how large–scale patters (e.g. NAO) are linked with windstorms and in this case their impact in particular. Therefore a composite study is performed to see whether or not years with, for example, positive NAO indices bear a higher risk of losses compared to years with negative NAO values. In a way this chapter ties all the other chapters back to the letter of Captain Richards in 1854: It uses a quantification (chapter 2) and large–scale drivers (chapter 3) on a regional basis (chapter 5) in order to make estimates about (monetary) losses (chapter 6). The report Captain Richards gave in his letter around 160 years ago is therefore more topical than ever...

Chapter 2

Severity quantification of European Winter Windstorms

"Measure what can be measured, and make measurable what cannot be measured."

– Galileo Galilei, Italian Polymath

Abstract

This paper introduces the Distribution–Independent Storm Severity Index (DI–SSI). The DI–SSI represents an approach to quantify the severity of exceptional surface wind speeds of large scale windstorms that is complementary to the Storm Severity Index (SSI) introduced by Leckebusch et al. (2008a). While the SSI approaches the extremeness of a storm from a meteorological and potential loss perspective, the DI–SSI defines the severity in a climatological perspective. The idea is to assign *equal* numbers to wind speeds of the *same extremeness* (e.g. the 99^{th} percentile). Especially in regions at the edge of the classical storm track the DI–SSI shows more equitable severity estimates, e.g. for the extra–tropical cyclone Klaus. In order to compare the indices, their relation with the North Atlantic Oscillation (NAO) is studied, which is one of the main large scale drivers for the intensity of European windstorms.

The following chapter is an edited and adapted version of the previously published article:

Walz, M. A., Kruschke, T., Rust, H. W., Ulbrich, U., & Leckebusch, G. C. (2017). Quantifying the extremity of windstorms for regions featuring infrequent events. Atmospheric Science Letters, 18(7), 315-322.

The initial idea of the publication that lead to the study for this chapter came from Tim Kruschke. Henning Rust initially suggested the use of the Generalised Pareto Distribution. Michael Walz carried out all of the analysis and created all the figures that lead to the publication. The manuscript was written by Michael Walz with Tim Kruschke and Gregor Leckebusch providing helpful comments which helped to improve the quality of the manuscript.

2.1 Introduction and Motivation

2.1.1 Background

Winter windstorms are among the biggest natural hazards occurring in the mid–latitudes causing human casualties as well as economic losses up to billions of Euros each year. According to SwissRe the winter storm Kyrill, which strongly affected Central Europe on January 18th, 19th 2007 caused an economic insured loss of about $6.1 billion and casualties of 54 people (SwissRe (2016)). An approach to objectively quantify the meteorological hazard is represented by the Storm Severity Index (SSI) introduced by Leckebusch et al. (2008a). The SSI is widely used (e.g. Osinski et al. (2016)) for assessing the severity of windstorms within the actuarial sector by linking extreme surface winds (i.e. exceedances of the 98^{th} percentile of local 6-hourly wind speeds) to potential loss on buildings. Furthermore, the 98^{th} percentile is used as a criterion for identifying extreme windstorms in a wind tracking algorithm by Leckebusch et al. (2008a) and further developed by Kruschke (2015). Equation 2.1.1 shows the mathematical definition of the SSI. The index t represents the time step, k represents the grid cell and A_k represents the area of the associated cell divided by a reference cell at the equator:

$$SSI_{T,K} = \sum_{t}^{T} \sum_{k}^{K} [(max(0, \frac{v_{k,t} - v_{98,k}}{v_{98,k}}))^3 * A_k] \tag{2.1.1}$$

The $v_{98,k}$ refers to the local 98^{th} percentile of the k–th grid cell which is the minimum wind speed at which damage on housing or nature is to be expected. This relationship was established based on real damage experience (Klawa and Ulbrich (2003)) which proved the assumption of Palutikof and Skellern (1991) who assumed

storm damages to occur at about 2% of all days.

For the further development of the index the focus will be on the Meteorological Contribution Γ to the SSI defined by Equation 2.1.2 which shifts and scales wind speeds by the 98^{th} percentile:

$$\Gamma = (v_{k,t} - v_{98,k})/v_{98,k} \tag{2.1.2}$$

Due to the division by the 98^{th} percentile, Γ^3 becomes a dimensionless number so that we cannot attribute a real physical meaning to it in contrast to v^3 which would represent the advection of kinetic energy.

2.1.2 Motivation for a supplementary severity index (DI–SSI)

Technically the SSI is an accumulation of wind speed excesses over a fixed quantile (percentile) for a given wind speed distribution; however the SSI does not take into account the shape of the distribution of these excesses, i.e. the tail behaviour. This becomes particularly obvious when estimating the SSI for areas with little storm occurrence, i.e. it results in very large SSI values. The top panel of Fig. 2.1.1 illustrates the reason for this effect: The two panels depict histograms of Γ (scaled and shifted wind speeds, cf. Eq. 2.1.2) of a grid cell south of Iceland (called Iceland hereafter) and on the Isle of Corsica. The coloured lines mark the 98^{th} (red) and the 99^{th} (green) percentile of the shifted and scaled distribution. The distance between those two percentiles is larger for Corsica. This shows that the contribution of local wind speed exceedances to the SSI can vary considerably between different regions thereby assigning too much weight to regions which are rarely affected by intense

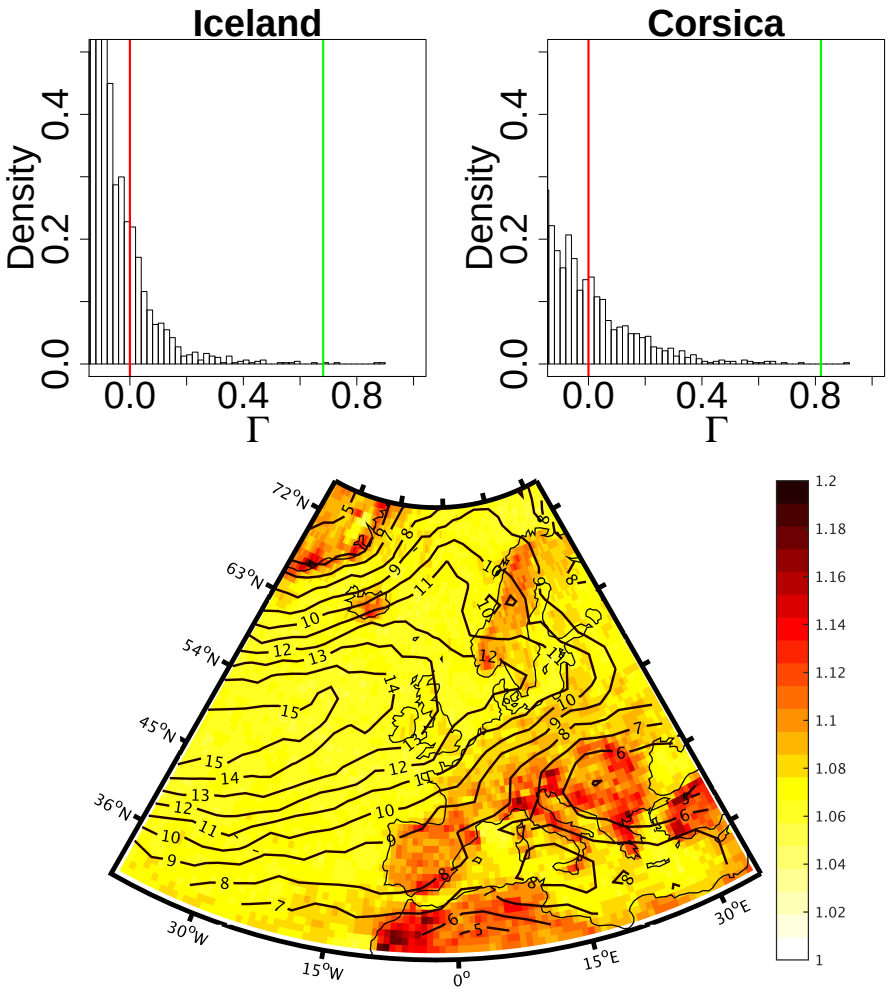


Figure 2.1.1: *Top: Histograms of the Meteorological contribution (Γ) as defined in Equation 2.1.2 for Iceland (left) and Corsica (right). The distribution of the events looks visibly different. The red line marks the 98^{th} percentile, thus the cut off threshold. The green line marks the 99^{th} percentile, thus it illuminates the larger difference between the two percentiles for Corsica in comparison with Iceland. Bottom: Quotient between the 99th and the 98^{th} percentile of local wind speeds for Europe in colour. Large values indicate a large difference between the percentiles. The contours depict the average track density per winter (ONDJFM; average number of tracks within a 500 km radius around a given grid point).The quotient is clearly larger in areas with a reduced windstorm frequency*

storms. Thus, the SSI cannot be used to directly compare the intensity of extreme windstorm events.

This study aims at creating a metric for the extremeness of a storm purely based on the shape of the tail of the local wind speed climatology. The desired outcome is to create a severity index which assigns the same number/value to storms with equal extremeness. For example, if there is a hypothetical storm over Iceland featuring only wind speeds of the local 99^{th} percentile it should be assigned the same value as a storm featuring wind speeds of the local 99^{th} percentile over the Mediterranean. The SSI in its current definition would be larger for the storm in the Mediterranean.

As the index should be a spatially and time integrated index as the SSI as well, this is trying to be achieved by creating parametric distributions for the tail of the distributions for every grid cell. Thus, for every grid cell we want to obtain an individual parametric distribution. We assume the new index to be independent of the nature of the tail distribution (e.g. heavy or light) as we account for every local distribution individually and obtain unique parameters based on the local wind climatology.

Due to being independent of the local distribution and its resemblance to the original SSI, the index is named Distribution–Independent Storm Severity Index (DI–SSI). The DI–SSI is particularly useful when comparing storms occurring in and outside of the main storm corridor. As an application, the two indices are correlated with the North Atlantic Oscillation (NAO), as it is currently recognised to be the most prominent driver of the inter annual variability of European storminess (e.g. Donat et al. (2010), Pinto et al. (2009) or Ulbrich and Christoph (1999)). Due to its para-

metric nature it is expected that the DI–SSI gives a more coherent and distinct link for areas outside of the classic storm track as it is a smoother function compared to the highly variable non-parametric SSI signal.

This paragraph shows in an exemplary way why the SSI tends to become large in areas with little storm activity:
By definition of the SSI, the red line in Figure 6.3.1 is equal to 0. The histogram for Iceland resembles a light tailed distribution whereas the histogram for Corsica shows features of a heavy tailed distribution. Accordingly, the gap between the 99^{th} and 98^{th} percentile is substantially different (0.68 for Iceland and 0.82 for Corsica). Note that due to the cubic construction of the SSI (Eq. 2.1.1), Γ is taken to the third power; for Corsica Γ is around 6 times larger than Γ^3 for Iceland (cf. Tab. 2.3.1). From a probabilistic perspective however both wind speeds are equally likely, i.e. have an equal occurrence probability. As the SSI is an accumulated number, the large differences in individual Γ contributions result in systematically larger SSI values in areas with heavy tailed wind speed distributions (c.f. examples for storms Klaus and Martin in Tab. 2.3.2).

Figure 2.1.1 (bottom panel) presents that problem on a larger scale: it depicts the quotient of the local 99^{th} and 98^{th} percentile and the average storm frequency per grid cell per extended winter season (i.e. how often on average a windstorm track is detected within a 500km radius of a particular grid cell). Klawa and Ulbrich (2003) calculated the same quotients for station data of wind speeds in various locations in Germany. Their conclusion was that the quotient was sufficiently homogeneous for the entire country. This assumption can be supported and confirmed by Fig. 2.1.1

(bottom). Values above 1.1, however, indicate areas in which storms are subject to large SSI values (for the reasons explained above), thus in particular Southern and Eastern Europe. These areas coincide with regions of little storminess over the winter period (less than 8-10 identified windstorm events per year).

2.2 Methods

2.2.1 Data and Event Identification

The wind speed data used for this work are taken from the ERA Interim reanalysis (Dee et al., 2011) which is managed by the European Centre for Medium Range Forecasts (ECMWF). The spectral resolution of ERA–Interim is T255 which corresponds to a grid cell of 0.7°x 0.7°at the equator. An objective wind-speed-based tracking algorithm (Leckebusch et al., 2008a; Kruschke, 2015) was applied to the 6–hourly 10-m wind field of the extended boreal winter period (October–March) in order to extract windstorm trajectories. ERA Interim has been frequently used in other ETC studies (Hodges et al., 2011). The NAO time series is obtained as the first principal component of a rotated EOF analysis of monthly (October–March) mean 700 hPa geopotential height anomalies for the North Atlantic domain (70°W-40°E, 30°N-80°N) as done by Hunter et al. (2016).

2.2.2 The DI–SSI

The derivation of the DI–SSI is based on the idea that excesses over a sufficiently large threshold can be well approximated by a Generalized Pareto Distribution (GPD). The approach of modelling excesses is one of the main concepts within Extreme Value Theory (EVT; see e.g. Coles, 2001). Modelling excesses of geo-

physical data with a GPD has been proposed in various other studies in connection with extreme precipitation (e.g., Vrac and Naveau, 2007 or Cooley et al., 2007), wind speeds (Kunz et al., 2010) and also SSI values (Donat et al., 2011; Held et al., 2013). The concept of the DI–SSI is to understand the numerator of the SSI equation (Eq. 2.1.1) as the exceedance of a threshold (i.e. the 98$^{\text{th}}$ percentile). In contrast to the common method of determining a threshold for the GPD, the threshold is fixed at the 98$^{\text{th}}$ percentile for every grid cell. The goodness of fit test provides satisfying results for this threshold (see below). A new variable is introduced to which the GPD is applied: $v^{\star}$ is defined as the random variable of the excess wind speeds over the local 98$^{\text{th}}$ percentile $v_{98,k}$ at grid cell k and time t:

$$v^{\star} = v_{k,t} - v_{98,k} \,|\, v_{l,t} > v_{98,k} \tag{2.2.1}$$

Estimating parameters of the GPD for $v^{\star}$ (using the ismev library in R; Heffernan et al., 2012) results in a pair of shape (ξ) and scale parameters (σ) for every grid cell. To get an idea of how well the GPD performs in describing $v^{\star}$ in the mid–latitudes, a Kolmogorov–Smirnov–test (ks–test) is used to assess the goodness of fit of the GPD distribution at every single grid point. Most grid cells over the North Atlantic and Europe do not show distances larger than the critical value D of the ks-test at the 5% significance level. Between 30°N and 70°N only 6% (2578 grid cells out of 43520) of all grid cells fail the test.

A potential spatial dependence of neighbouring grid cells is neglected as each grid cell is considered as an individual contributor to the DI–SSI, although spatial dependence would potentially increase the amount of rejected cells. Being aware of the weaknesses of the ks-test when distributional parameters are estimated from

the sample and the multiple-testing setting, we still consider this test as evidence that a GPD represents a sufficiently good model of $v^\star$ in our region of interest. To avoid the problem with estimated distributional parameters, one could simulate the distribution of the test-statistic under the Null for every grid point. However, we consider this as too costly for the scope of this study here.

Analogous to the equiprobability transformation to yield the Standardized Precipitation Index (SPI; cf. Lloyd-Hughes and Saunders, 2002, their Figure 1 shown here as Figure 2.2.1), the GPD fitted cumulative probability distribution of $v^\star$ is transformed into a standard exponential distribution. This equiprobability transformation is performed to map the empirical cumulative density function (ECDF) of each individual GPD distribution onto comparable values. Technically the choice

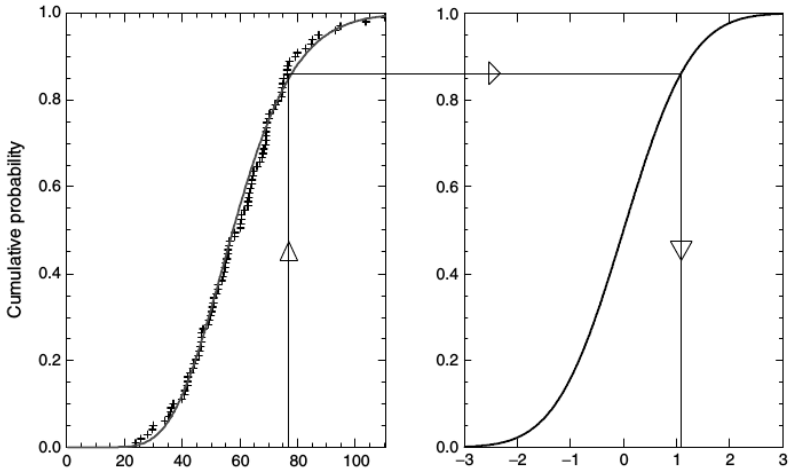


Figure 2.2.1: *This figure represents an exemplary equiprobabilty transformation (taken from Lloyd-Hughes and Saunders, 2002). The cumulative probability of the fitted values (left side) are mapped onto the probabilities of the new distribution. In the case of the DI–SSI this would be the standard exponential distribution.*

of distribution to transfer the GPD to is arbitrary as it only determines the range of the index values. As the GPD is part of the exponential family we choose a stan-

dard exponential here (rate $\lambda = 1$). Equation 2.2.2a defines the transformed value x: by equating the GPD (on grid cell level) with the standard exponential probability distribution for x. The equation is then solved for x. The resulting Eq. 2.2.2b gives the contribution to the DI–SSI of a single grid cell where ξ represents the shape and σ the scale parameter of the GPD distribution.

$$1 - (1 + \frac{\xi v^\star}{\tilde{\sigma}})^{\frac{-1}{\xi}} = 1 - e^{-\lambda x}; \; \lambda = 1 \tag{2.2.2a}$$

$$\frac{1}{\xi} ln(1 + \frac{\xi v^\star}{\tilde{\sigma}}) = x \tag{2.2.2b}$$

The definition of the (integral) DI–SSI in turn is the result of the summation of the contributions x over the entire footprint of a respective storm, equivalent to the definition of the SSI (compare Eq.2.1.1 and Eq. 2.2.3).

$$DI - SSI_{T,K} = \sum_{t}^{T} \sum_{k}^{K} [\frac{1}{\xi} ln(1 + (\xi v^\star)/\sigma) A_k] \tag{2.2.3}$$

2.3 DI–SSI in practice and in comparison the SSI

Table 2.3.1: *Meteorological and DI–SSI contributions of the two example grid cells. Note that the contribution of the grid cell in Corsica is more than five times larger, although the wind is of the same extremeness in both cases.*

	Γ^3 in a single grid cell for a surface wind equal to the 99$^{\text{th}}$percentile	DI–SSI contribution in a single grid cell for a surface wind equal to the 99$^{\text{th}}$ percentile
Iceland	$2.05 \cdot 10^{-4}$	0.71
Corsica	$1.24 \cdot 10^{-3}$	0.67
Theoretical value	–	0.69

In order to compare the SSI contributions Γ to its DI–SSI equivalents (x in Eq. 2.2.2b), both were calculated for the grid cells described in Chapter 2.1.2. As opposed to

the SSI contribution for that particular grid cell, which differs by a factor of almost 20, the DI–SSI contribution is almost equal for the two wind speeds (see Tab. 2.3.1). The 99$^{\text{th}}$ percentile of the original wind speed distribution V_k at grid cell k is equal to the 50th percentile of $v^{\star}$ (as only wind speeds above the 98$^{\text{th}}$ percentile are considered). By definition of the standard exponential distribution its 50$^{\text{th}}$ percentile (median) is equal to $ln2 = 0.69$ (theoretical value in 2.3.1). Thus a wind speed v_k equal to the 99$^{\text{th}}$ percentile results in a DI–SSI contribution x of 0.69.

Table 2.3.2: *Integral SSI and DI-SSI values for some prominent European windstorms. The rank of severity for the respective index is denoted in brackets. Note that storms which occurred outside of the main storm tracks feature relatively large SSI values (e.g. Klaus, Martin, Xynthia, Torsten) compared to the ones within the main storm track (Daria or Jeanette). This applies especially for the SSI/DI–SSI values per time step. The largest discrepancy in terms of rank of the integral values is observed for Martin and for Vivian/Klaus for time step based values.*

Storm	*Date*	*Integral SSI value*	*Integral DI–SSI value*	*SSI per time step*	*DI–SSI per time step*	*References*
Daria	23–26 January 1990	26.69[7]	1940.20[4]	2.05[8]	149.25[5]	Heming (1990)
Vivan	25–28 February 1990	58.52[2]	4126.34[1]	3.90[4]	275.10[1]	McCallum and Norris (1990)
Anatol	2–4 December 1999	23.57[8]	1565.67[6]	1.81[9]	120.44[8]	Ulbrich et al. (2001)
Martin	26–28 December 1999	43.81[4]	1435.09[8]	5.48[2]	179.39[3]	Ulbrich et al. (2001)
Torsten	10–13 November 2001	15.94[9]	789.95[9]	2.66[6]	131.66[7]	Tripoli et al. (2005)
Jeanette	25–31 October 2002	32.53[6]	1576.27[5]	2.32[7]	112.60[9]	Parton et al. (2009)
Kyrill	15–24 January 2007	53.03[3]	2439.57[2]	4.08[3]	187.66[2]	Fink et al. (2009)
Klaus	23–28 January 2009	74.30[1]	2117.52[3]	5.72[1]	162.89[4]	Liberato et al. (2011)}
Xynthia	26 February–7 March 2010	37.92[5]	1459.10[7]	3.45[5]	132.65[6]	Lumbroso and Vinet (2011)

Table 2.3.2 presents integral values of the SSI and DI–SSI for some of the most prominent European windstorms. As expected storms that occurred on the edges of the classical storm track yield comparatively large SSI values. One of the most striking example is represented by windstorm Klaus (Liberato et al., 2011) whose SSI value is almost three times as large as the respective value for windstorm Daria (Heming, 1990), whereas their DI–SSI values are almost of the same magnitude.

Similar conclusions can be drawn from the storms Klaus and Kyrill (Fink et al., 2009): The DI–SSI is similar for both events; however the SSI is about 1.5 times

larger for Klaus. Thus, judging from the SSI it appears that storm Klaus was far more intense than both Daria and Kyrill. The different assessment of severity for storms in different climatic background conditions is even more striking when comparing average SSI/DI–SSI values per time step. Klaus and Martin (Ulbrich et al., 2001) which follow similar tracks across Southern and Central Europe exhibit the largest SSI values per time step whereas the largest DI–SSI per time step can be identified for Vivian and Kyrill. Daria, Klaus and Vivian show the largest difference in rank if assessed by the average value per time step.

The largest DI–SSI is associated with the storm Vivian (McCallum and Norris, 1990) which ranks second with regard to the SSI ranking. The large magnitude of the DI–SSI can potentially be explained by very extreme winds observed over the Atlantic Ocean (cf. Figure A.1.1). As shown for storm Klaus in Chapter 2.2.2, the DI–SSI contributions over the sea are considerably larger than for the SSI. Thus, a storm with extreme surface winds over the sea is subject to high DI–SSI values as the DI–SSI is purely based on the extremeness of wind speeds without any potential impact consideration. The biggest discrepancy between the respective rankings for the integral values of the storms is observed for storm Martin (4^{th} compared to 8^{th}) which is in line with the arguments for storm Klaus.

An application of both indices is shown in Fig. 2.3.1 (left panel) where the correlation of the annual storm intensity (annual sum of all SSI/DI–SSI contributions within a 500 km radius around a grid cell) and the annual storm occurrence per grid cell is presented. Hunter et al. (2016) showed a similar figure (their Figure 4(a)) using the vorticity as a severity metric. The coherent area of significant values over

Scandinavia is smaller compared to their results. The overall pattern looks fairly similar though, with most of Scandinavia showing positive correlations, implying that seasons with more storms also feature more intense storms. Especially for the

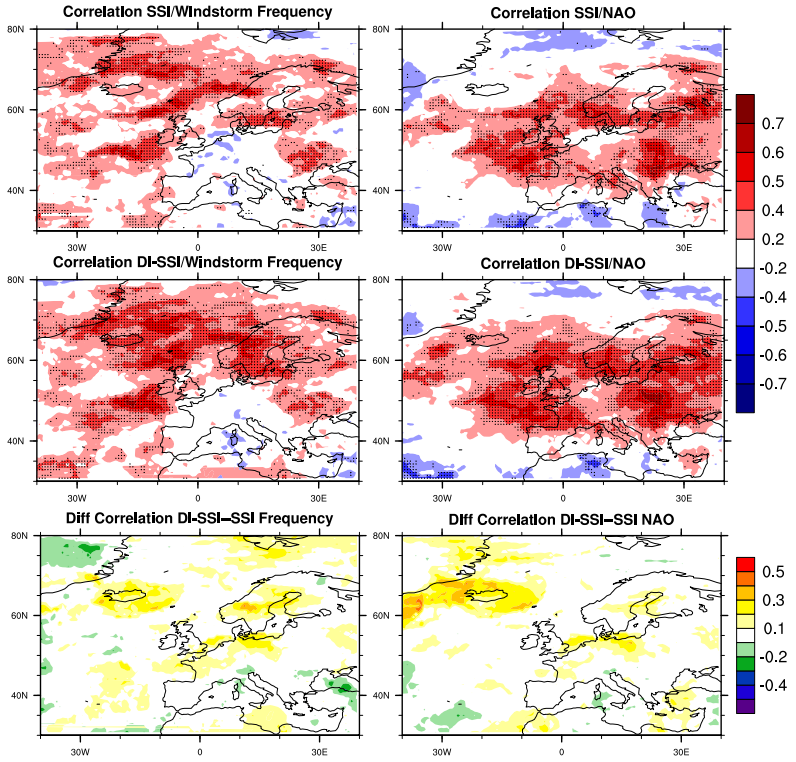


Figure 2.3.1: *Left panel: Correlation between the storm frequency and storm intensity (SSI on the top; DI–SSI in the middle) for each grid cell. Correlation coefficients significant at the 5% level are stippled. There is a significant link between more storms and more intense storms for much of the North Atlantic and Scandinavia. Right panel: Correlation coefficients between the yearly NAO time series and the yearly windstorm intensity on grid cell level (SSI on the top; DI–SSI in the middle). Again correlation coefficients significant at the 5% level are tagged. Bottom row: Differences between the respective correlations. Positive values represent areas where the correlation for the DI–SSI is larger compared to the SSI. This is the case for most of the North Atlantic Domain.*

DI–SSI (middle-left panel) there is a large area of significant correlation between occurrence and intensity southwest of the British Isles that is not visible in their

figure. This indicates the enhanced capability of the DI–SSI to characterise intense and unusual wind speeds not only over land but also over the sea compared to using vorticity as a severity metric. This is in line with the large DI–SSI value for windstorm Vivian for the DI–SSI is capable of quantifying extreme surface winds regardless of their occurrence. This is also supported by the difference between the two correlations shown in the bottom-left panel of Fig. 2.3.1 as most areas over the Central Atlantic are positive, thus denoting larger DI–SSI correlations.

2.3.1 SSI and DI–SSI compared for a European storm example

Figure 2.3.2 serves as an example of how the previously discussed differences between the SSI and the DI–SSI arise: The figure shows a snapshot of the footprint of storm Klaus (Liberato et al., 2011) and the footprint of the entire storm in the bottom right panel. The overall footprint of the storm looks exactly the same by definition as the local 98^{th} percentile is used as a detection criterion in the storm tracking algorithm. The geographical intensity distribution however is different for the two indices (both indices are standardised for comparison). Whereas the SSI has its largest contributions over land on the northern coast of the Iberian Peninsula, the DI–SSI has in fact its largest contributions over the sea just north of the northern coast of Spain. This area coincides with the area of the most extreme wind speeds. This difference becomes more obvious when looking at the differences of the contributions of both indices. The coast line of the Iberian Peninsula represents an almost perfect segregation between negative and positive differences. This is according to the expectation regarding the features of the SSI and DI–SSI.

The SSI can be used well to assess the potential damage to infrastructure, however judging from this figure it would seem that the wind speeds over the Atlantic do not have the same exceedance probability as they have over land. The DI–SSI draws a different picture: Albeit still showing large values over land, the more extreme values are apparent over the ocean indicating that the wind speeds in that area were even more exceptional with regard to their climatological wind speed distribution. This supports the arguments regarding the large DI–SSI for the storm Vivian in Chapter 2.2.2.

2.4 Intensity indices in connection with the NAO

A more quantitative comparison is supplied in the right panel of Fig. 2.3.1. These two figures show the correlation coefficients between the annual winter NAO time series and the annual intensity time series per grid cell for the SSI and DI–SSI respectively. Grid cells with a correlation coefficient significantly different from zero at the 5% level are stippled. This correlation does not necessarily prove any physical evidence, however it indicates that the correlation was unlikely if the null hypothesis was true. Considering this fact, both maps show a significant link between the NAO and the intensity of European windstorms for most of Europe. However, overall there are more significant grid cells for the correlation using the DI–SSI compared to the SSI. This applies especially to large parts of southwest France, parts of Northern Africa and some areas in northeast Europe, thus regions which are affected less frequently by large scale windstorms.

The largest difference in correlation is observed south of Greenland and around Iceland. According to the bottom panel of Fig 2.1.1 these areas are also on the

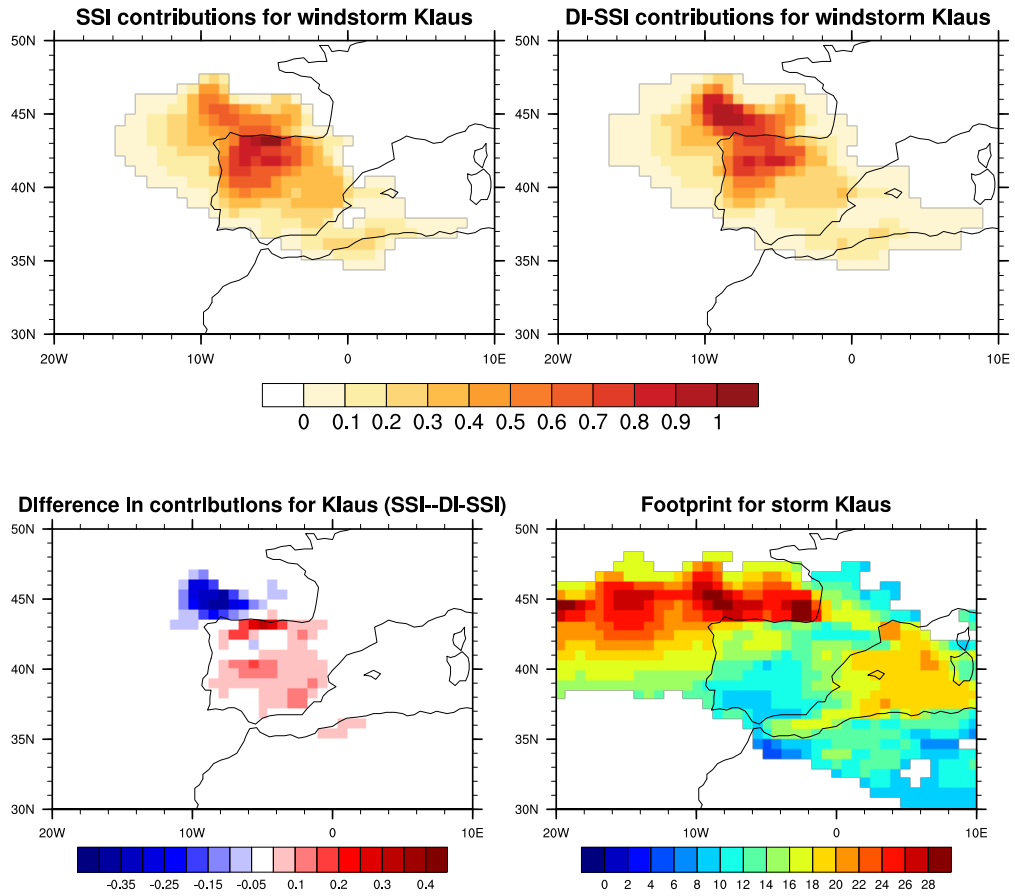


Figure 2.3.2: *Footprint of storm Klaus on 24 January 2009 with SSI contributions shown on the top left, the DI–SSI contributions on the top right and the differences between the both values on the bottom left. The footprint of maximum wind speeds* $(m \cdot s^{-1})$ *for the entire storm is shown in the bottom right panel. Both SSI and DI–SSI were standardized for comparison reasons. Positive values indicate grid cells with larger SSI contributions value; negative values indicate larger DI–SSI contributions. There is a distinct separation represented by the coast line of the northern and western coast of the Iberian Peninsula.*

edge of the storm track. This is another indication showing that the DI–SSI is a suitable metric to quantify extreme windstorm events occurring outside the main storm track. The correlation pattern within the main storm track (central North

Atlantic) is almost equal for both indices This supports the expectation that they behave fairly similar given the amount of storms per grid cell is sufficiently large. Thus, the DI–SSI is a useful metric to represent the extremeness of wind speeds both in areas with little annual storm activity and also in areas with increased storminess.

2.5 Summary and Discussion

This study introduces the Distribution independent Storm Severity Index (DI–SSI): It serves as a quantification of extreme windstorms, especially for those occurring outside of the main storm tracks. Due to strongly diverse wind climatologies in different regions, the actual wind speed is an improper metric for the assessment of extremeness. A widely used index, especially in the impact community is the SSI developed by Leckebusch et al. (2008a). The SSI is a metric that relates extreme winds to their potential damage on housing or infrastructure, whereas the newly introduced DI–SSI assesses the severity of exceptional wind speeds based on their occurrence probability: Wind speeds with the same occurrence probabilities contribute in an equal way to the DI-SSI index regard less of the location of the storm event. In case of the SSI contributions from areas with heavy–tailed distributions would be weighted more.

SSI/DI–SSI values are presented for nine prominent European winter storms. These values reveal the difference between the two indices. The largest SSI values arise for storms occurring on the edges of the storm track (Klaus, Martin), whereas the DI–SSI ranks storm Vivian and Kyrill as the most severe events.

In connection with the NAO index, the DI–SSI time series shows a more coherent

area of significant correlation over Southwest Europe and also the Baltic Sea area compared to the SSI. This proves the capability of the DI–SSI to assess the severity of extreme winds both inside and outside of the main storm track. A larger area of correlation is also apparent for the correlation between frequency and intensity. The overall pattern of correlation for both indices is in agreement with Hunter et al. (2016). The results imply that especially within the main storm track in the North Atlantic and for most parts of Scandinavia, seasons with many storms also tend to feature more intense storms. This is in accordance with Vitolo et al. (2009) who found that serially clustered seasons are likely to spawn more intense storms. Technically, the SSI is easier to compute than the DI–SSI for it only requires wind speed data on grid cell level and no fitting of a statistical model. The DI–SSI requires more processing of the data, however it is a useful additional tool to assess the severity of storms/extreme winds regardless of their geographic occurrence.

Chapter 3

Large scale drivers and temporal variability of Windstorms

"Essentially, all models are wrong, but some are useful."

– George E.P. Box, British Statistician

Abstract

Winter windstorms are known to be amongst the most dangerous and loss intensive natural hazards in Europe. In order to gain a better understanding of their variability and driving mechanisms, this study analyses the temporal variability which is often referred to as serial or seasonal clustering. This is realised by developing a statistical model relating the winter storm counts to known teleconnection patterns affecting European weather and climate conditions (e.g., North Atlantic Oscillation (NAO), Scandinavian Pattern (SCA), etc.). The statistical model is developed via a stepwise Poisson regression approach that is applied to windstorm counts and large scale indices retrieved from the ERA-20C reanalysis. Significant large scale drivers accountable for the inter-annual variability of storms for several European regions are identified and compared. In addition to the SCA and the NAO which are found to be the essential drivers for most areas within the European domain, other teleconnections (e.g. East Atlantic Pattern) are found to be more significant for the inter-annual variability in certain regions. Furthermore, the statistical model allows an estimation of the expected number of storms per winter season and also whether a season has the characteristic of being what we define an active or inactive season. The statistical model reveals high skill particularly over British Isles and Central Europe, however even for regions with less frequent storm events (e.g., Southern and Eastern Europe) the model shows adequate positive skill. This feature could be of specific interest for the actuarial sector.

The following chapter is an edited version of the previously published article: *Walz MA, Befort DJ, Kirchner-Bossi NO, Ulbrich U, Leckebusch GC. Modelling serial clustering and inter-annual variability of European winter windstorms based on large-scale drivers. Int J Climatol. 2018;38:3044-3057.*

This study is the result of a project with the Risk Prediction Initiative (`http://rpi.bios.edu/`) which was won by Gregor Leckebusch and Uwe Ulbrich. Nicolas Kirchner-Bossi carried out some initial data processing (calculation of the monthly indices and tracking of the storm events). Michael Walz and Daniel Befort developed the statistical model used in this study. Michael Walz carried out the analysis and produced all the figures which ended up in the publication except for Figure 3.3.1 which was produced by Nicolas Kirchner-Bossi and Figure A.2.1 which was created by Daniel Befort. The manuscript was written by Michael Walz with Daniel Befort and Gregor Leckebusch contributing to the writing.

3.1 Introduction

Winter windstorms embody a prominent feature of the European climate. They are often accompanied by severe surface winds that can result in extensive socio-economic losses. More than half of the insured loss caused by natural hazards in Central Europe emanates from extreme winter windstorms (MunichRe, 2007). Various studies have discussed the potential forcing factors influencing the inter-annual storm variability expressed in increased or decreased numbers of potentially destructive cyclone systems per season. An atmospheric variability pattern that is frequently considered in connection with the occurrence of winter storms is the North Atlantic Oscillation (NAO; Walker, 1928, Hurrell, 1995). The association between the NAO and European storminess has been extensively examined in previous studies. Pinto et al. (2009) showed that a positive NAO phase is in favour of growth conditions of storms compared to a negative NAO phase. Likewise, Donat et al. (2010) found that more than 20% of storm days occur within a strongly positive NAO phase even though this period is found on less than 7% of all days.

Other atmospheric teleconnections that have been detected as driving factors for the variability of the European winter climate include the East Atlantic/West Russia pattern (EA/WR, e.g. Lim, 2015) or the Scandinavian pattern (SCA, e.g. Bueh and Nakamura, 2007). Seierstad et al. (2007) related extra-tropical storminess, defined as monthly mean variance of high-pass filtered sea level pressure, to large scale patterns by using a Gamma regression. They showed that five teleconnection patterns are significant at the 5% level with regard to explaining the inter-annual variabil-

ity: NAO, SCA, East Atlantic Pattern (EA), EA/WR and the Polar Pattern (POL).

More recently, Hunter et al. (2016) found a significant correlation between cyclone counts in Scandinavia and the SCA index. In terms of a physical link between large scale patterns and European storminess Woolings and Blackburn (2012) could detect a link between the NAO and the EA on the location and strength of the North Atlantic Jet Stream which is in turn responsible for increased or decreased storminess during the European winter. This link works both ways as the Jet Stream also has an impact on the NAO and the EA and vice versa.

Other studies suggest that variations in the sea surface temperature (SST) in the North-Atlantic also act as an important driver of variability of European storminess. Periods of high decadal storm activity were identified to be preceded by a phase of a weak North-Atlantic meridional overturning circulation (MOC) by Nissen et al. (2014). This is due to a distinctive change in the mixed ocean layer heat content (OHC).

Saunders and Qian (2002) and Czaja and Frankignoul (2002, 1999) found a Horseshoe-shaped anomaly pattern of North Atlantic SST in summer and autumn that exhibited a strong link to the NAO in the subsequent winter. These findings could be supported by Renggli (2011), who showed that a horseshoe pattern in autumn is linked to windstorm frequency in the subsequent winter. One of the factors identified for the extreme storm season of the winter 2013/14 by Wild et al. (2015), is the meridional temperature gradient between the North-American continent and the SSTs in the West-Atlantic. According to their study there is a significant correlation

between this temperature gradient and windstorm occurrences over the Eastern Atlantic, the Iberian Peninsula and the south west of the British Isles.

Understanding the serial clustering of winter windstorms is a key component in comprehending their inter-annual variability. It is of particular interest within the actuarial industry as temporal clustering is responsible for large accumulated losses over an entire storm season. Seasonal clustering has been investigated statistically by Mailier et al. (2006), Vitolo et al. (2009) and Pinto et al. (2013). Their studies reveal the overall pattern of cyclone clustering (over-dispersion) on both sides and downstream of the North Atlantic storm track, while under-dispersion is found around the entrance of the storm track, hence close to Newfoundland. Recently, Pinto et al. (2016) showed that the statistical features of serial clustering and the influence of the NAO on serial clustering are independent from the storm/cyclone tracking algorithm used to identify the events.

The common statistical definition of seasonal clustering is to examine the deviation of a windstorm count time series from the Poisson distribution which features equal mean and variance. Thus a natural approach is to verify to what extent the annual storm count time series follows a random point process or if the occurrences of storms are of a more systematic nature resulting in unequal mean and variance. This is realised by a dispersion statistic (variance to mean ratio) which is used to quantify the deviation from the random Poisson process. Vitolo et al. (2009), for example, found that monthly clustering is linked with the intensity of the storms. Thus, clustered seasons are likely to feature more intense windstorms than on average. Additionally, they were able to reproduce the dispersion statistic by using a

Poisson regression model in connection with some of the large scale teleconnection patterns discussed above. Mailier et al. (2006) came to a similar result as they also identified 5 teleconnection patterns (NAO, EA, SCA, EA/WR, POL) that have a significant impact on the inter-annual variability of cyclone counts for Europe, noting however that only the NAO by itself is not capable of explaining the entire variance.

Economou et al. (2015) looked at the capability of 17 CMIP5 models to capture the clustering which is observed in the reanalysis data. In particular in the northern and the southern part of the Atlantic storm track and over Western Europe they were able to show the over-dispersion of extreme wind storms counts. Additionally they found that the variability of the NAO explains a large part of this over-dispersion in the historical runs for the same areas. The question that arises from the results of these previous studies is, if serial clustering can be explained and modelled by a statistical model, thus if large scale drivers can be directly utilized to estimate the amount of windstorms per season and thus the serial clustering of the overall time series. To examine these questions this paper is aiming at answering two central questions:

(a) Looking at different predefined regions in Europe: What are the main drivers responsible for serial clustering in a particular region?

(b) After having identified several prominent drivers: Where is the main area of influence of these predominant European drivers on the inter-annual variability of windstorms?

Question (a) is addressing the impact perspective of serial clustering as the regions

are roughly in accordance with areas used within the actuarial sector whereas question (b) is examining this topic from a more physical angle: As the defined regions within the actuarial sector might split some of the areas of influence, our goal is to exactly locate the zone of influence of the dominant large scale drivers on the inter-annual variability of windstorms. In order to precisely allocate drivers to different areas in Europe we are answering question (b) on grid cell level, thereby creating a "map of drivers" for the European domain.

The intention of this study is to gain a better understanding and to quantify the inter-annual variability of winter windstorm occurrence over Europe. This is achieved by investigating the influence of previously discussed teleconnections as well as testing further potential large scale drivers that have not been examined as thoroughly with regard to winter windstorms. Due to its importance for e.g. the insurance sector special focus is put on the link between these large-scale drivers and serial clustering.

As the reanalysis product ECMWF ERA-20C (Poli et al., 2016) is relatively new, this investigation has never been carried out on the time scale of an entire century. We are aware of the potential handicap of ERA-20C due to the lack of constraints for the reanalysis especially in the first half of the 20th century. As the scope of this study is solely the understanding of the physical drivers of windstorms, however, we leave the assessment and validation of this reanalysis product to further studies and also refer the reader to Befort et al. (2016) who compared the windstorm climatologies in ERA-20C and NOAA 20CR (Compo et al., 2011).

Additionally all previous studies on serial clustering of winter storms (e.g. Mailier et al., 2006 or Vitolo et al., 2009) utilised mean sea level pressure (MSLP) based tracking schemes to identify storm events whereas this study applies a wind based tracking algorithm. Additionally, the map of drivers could provide a useful overview over the spatial distribution of large scale drivers across the European domain.

3.2 Data

The ECMWF reanalysis ERA-20C (Poli et al., 2016) was used to identify windstorm events (featuring trajectories and footprints) for the core winter season December-February (DJF) during the past century using a wind tracking algorithm which is based on the local exceedance of the 98th percentile of wind speeds (Befort et al., 2016, Kruschke, 2015, Leckebusch et al., 2008a). The identified windstorm trajectories are used to determine track densities of annual windstorm counts. This is done either as simply counting annual windstorms passing through one of the seven regions (cf. Figure 3.2.1) or, for the grid cell approach, as annual windstorm counts passing through a 500km radius around each 1° grid cell. Stalling or slow moving systems are only counted once for a respective grid cell ensuring a correct count number of storms. The respective time series of counts (either for the region or for the grid cell) is used as the predictant for the regression approach.

The pool of potential large scale drivers (predictors) contains 20 normalized and standardised (by the standard deviation) index time series, including the 10 leading rotated EOFs (Figure A.2.1 in the Appendix) of 700 hPa geopotential height anomalies over the northern hemisphere (calculation according to <http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml>), the station based NAO index

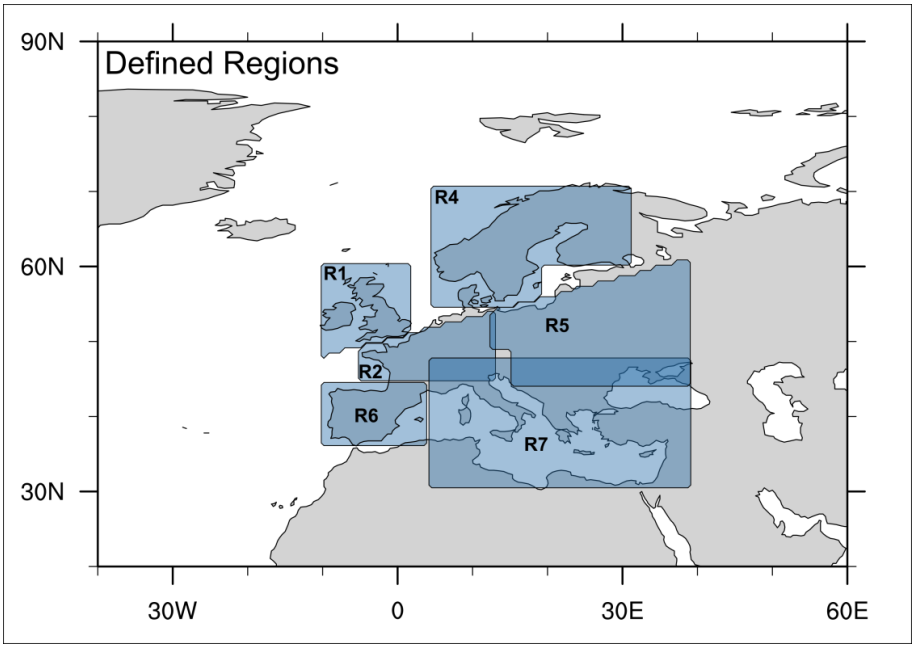


Figure 3.2.1: *Defined European sub-regions. R1: British Isles, R2: Central Europe, R3: British Isles + Central Europe(R1+R2), R4: Scandinavia, R5: Eastern Europe, R6: Iberia, R7: Mediterranean.*

and several SST related indices. A varimax rotation method was used for the rotation of the EOFs (Kaiser, 1958). The advantage of the rotated EOFs over the non–rotated ones is the fact that they are more easily interpretable as loadings of medium magnitude are either increased or decreased so that the variance is maximised. Hence the name varimax rotation. In contrast to non–rotated EOF time series, rotated ones can be correlated. This is not a problem however, as we perform a check for co–linearity while constructing the model selection (see VIF below). Table 3.2.1 gives an overview of the nomenclature of the indices as used in the paper.

The complete list plus the definition used to derive each index can be found in Table A.5.1 in the Appendix. The Quasi-Biennial Oscillation (QBO) time series in the pool of potential drivers was kindly provided by Brönnimann et al. (2007) and

Table 3.2.1: *Large scale indices nomenclature as used in the paper.*

Index name	*Long name*
QBO30	Quasi-Biennial Oscillation (30 hPa)
QBO70	Quasi-Biennial Oscillation (70 hPa)
AMO	Atlantic Meridional Oscillation
HSI	Horse-Shoe-Index
SSTS	Southern Box of HSI
Tdif.Nam	Temperature difference North America – West Atlantic
W.Atl T	West Atlantic SST
NINO3.4	Nino 3.4 index
NAO.Is.Li	Station Based NAO index
PDO	Pacific Decadal Oscillation
West Pac	West Pacific Pattern (EOF)
PNA	Pacific-North American pattern (EOF)
EOF10	West Pacific pattern II (EOF)
EA.WR	East Atlantic/West Russia pattern (EOF)
EA	East Atlantic pattern (EOF)
SCA	Scandinavian Pattern (EOF)
TNH	Tropical Northern hemisphere (EOF)
EP.NP	East Pacific/North Pacific pattern (EOF)
POL	Polar index (POL)
Sea Ice	Northern Hemispheric sea ice cover

is thus not calculated from the ERA-20C reanalysis. The sea ice index time series is constructed by normalising and standardising the northern hemispheric sea ice extent between 40°N and 90°N during the core winter season (DJF) from ERA-20C.

The 20 predictors were tested for co-linearity by calculating the variance inflation factor (VIF, e.g. O'brien, 2007) for each potential predictor. The VIF reflects the proportion of variance in one predictor that is explained by all the other predictors in the model. In practice the VIF is calculated for each variable as the reciprocal inverse of the coefficient of determination. A VIF of 1 would indicate no co-linearity,

whereas larger values suggest increasing co–linearity between the predictors. The common threshold used as a cut off value is a VIF of 10 (e.g. Hair Jr et al., 1995). Based on the VIF criterion and the skill of the model, the NAO derived from EOF analysis was discarded as the station–based index yielded higher correlations with annual storm counts.

3.3 Development of the statistical Poisson model

This section shall provide an overview of the two perspectives used in this study: Firstly the regional, more impact based statistical model is introduced. Ensuing, the grid cell based Poisson model approach is described in detail. In order to investigate regional characteristics of windstorm clustering in Europe, a statistical model is developed for 7 different European sub-regions. These regions are defined by both meteorological and socioeconomic criteria which are widely used in the actuarial sector (cf. Figure 3.2.1). As described in Chapter 3.2, the amount of storms per core winter season (DJF) within the limits of a region is counted and used as the predictant in the stepwise regression approach.

The procedure of finding the main large scale drivers for every region is based on a stepwise AIC (An Information Theoretic Criterion; Akaike, 1974) approach. The AIC is an information-theory-based (e.g. Jaynes, 1957) estimate of how much information is lost if using a statistical model instead of the actual physical relation. Thus, the AIC can be used as a tool for model selection when different models are available. Part of the selection process is the trade-off between the goodness of fit of the model and the complexity of the model, as the number of parameters to be estimated k, as well as the maximum of the likelihood function L for estimating the

regression coefficients are part of the AIC score (cf. Equation 3.3.1).

$$AIC = 2k - 2\,ln\,L \tag{3.3.1}$$

The model yielding the smallest AIC is declared as the chosen/winning model. Unlike other model selection criteria, e.g. the F-test, the AIC score does not provide any evidence about the absolute quality of the model. Thus, in case all available models are poor, there is still a winning model albeit it being of poor quality. To account for that the chosen predictors of the best AIC model are tested for statistical significance using a Wald - or χ^2 –test (e.g. Agresti, 2003). Only predictors significant at the 5% level are included in the final model.

The selected predictors for every region are further used for to fit a Poisson Generalised Linear Model (GLM). As the Poisson GLM is the recognised model for modelling count data (e.g. Vitolo et al., 2009) we consider it to be the natural choice. Due to a significant trend in windstorm counts identified in Befort et al. (2016), a linear time trend coefficient is added to the Poisson GLM. The model definition is given by Equation 3.3.2a:

$y(t)$ represents the number of storm counts in season t, $\lambda(t)$ represents the Poisson mean, $x_1(t)$ represents the time (108 DJF seasons) and the $x_i(t)$ represent the previously selected large scale drivers. The coefficients for the respective predictors are

labelled with β_i.

$$log(\lambda(t)) = \beta_0 + \beta_1 x_1(t) + \sum_{i=2}^{N} (\beta_i x_i(t)) \tag{3.3.2a}$$

The model is intended to capture the observed serial clustering given by the overall dispersion statistic over these 7 regions. The clustering dispersion score (Mailier et al., 2006) estimates how well the model is capable of reproducing the clustering of the entire time series. It is assessed through the index of dispersion (D), also referred to as the Variance-to-mean ratio, where the mean and the variance are simply calculated from the respective time series.

$$D = \frac{\sigma^2}{\mu} - 1 \tag{3.3.3}$$

The mean and variance are equal for an ideally Poisson distributed variable (leading to $D = 0$), thus the occurrence of an event is independent (in a statistical sense) of the timing of the previous event. This implies that a deviation of D from 0 suggests some kind of a serial dependence of successive statistical events. In terms of windstorms this indicates that successive storm seasons might not be independent of each other. This deviation of D from zero is the definition/quantification of irregularity, or in terms of windstorms, clustering. If the index of dispersion D is larger than 0 the time series is considered to be over–dispersed (thus following a negative binomial distribution). Referring to the annual storm counts this implies that the occurrence of storms happens in clusters for individual years. The model for all seven different regions is compared via the clustering dispersion bias ($D_{model} - D_{obs}$) thus, the difference between the modelled dispersion statistic and the actually ob-

served dispersion statistic D, which is calculated from ERA-20C directly.

As the dispersion statistic D by itself conveys no information about the clustering that happens within a particular season, we have decided to assess the regularity/irregularity of a particular season by a metric which we name active or inactive season. A season is considered an active season (AS) if it spawns more windstorms than one standard deviation above the long-term mean. Equivalently, a season is considered an inactive season (IAS) if it features fewer windstorms than one standard deviation below the mean.

$$AS_{r,t} = Cts_{r,t} \,|\, Cts_{r,t} > \mu_r + \sigma_r \tag{3.3.4a}$$

$$IAS_{r,t} = Cts_{r,t} \,|\, Cts_{r,t} < \mu_r - \sigma_r \tag{3.3.4b}$$

Here µr represents the mean of the long term storm count time series in region r and σ_r represents its standard deviation. $Cts_{r,t}$ represents the windstorm count for DJF t in region r. The amount of IAS/AS is compared for the observations and the model. This second metric represents an aspect of serial clustering that is more tailored towards the actuarial community as it conveys information on whether or not a season features above or below average storm counts, thus an increased or decreased likelihood of two or more storms in succession.

To examine in how far the Poisson GLM is able to reproduce these active and inactive seasons we define a hit rate (HR): The hit rate is defined as the quotient of the count of correctly predicted and observed active or inactive seasons over the number of observed active/inactive seasons within the entire time series. This metric is used

to assess the model's ability to capture active and inactive season. The HR will be given as a percentage of correctly predicted active seasons. Thus, regardless of the actual count of storms, a season is considered to be predicted correctly if it exceeds the AS or IAS thresholds.

All comparisons are made to the observed windstorm counts derived from ERA-20C. In order to account for potential over fitting a 10-fold cross validation method (Krstajic et al., 2014) is applied to the statistical model for all seven regions. The skill scores presented in the tables in Chapter 3.4.1 are calculated using the cross-validated model.

In order to examine the physical perspective of the influence of large scale drivers on the inter-annual variability of windstorms (represented by track density) on grid cell level, we implemented an independent Poisson GLM approach for every grid cell using five predominant drivers which are partly taken from literature (e.g. Mailier et al., 2006 or Vitolo et al., 2009) and partly from results of the impact based statistical model. As the intention is to comprehend annual windstorm count on grid box level, a Poisson regression model appears as the natural choice again (cf. Equation 3.3.2a).

An annual track density of windstorms per core winter season (DJF) is calculated on a 1° grid cell level for the North Atlantic domain (40°W-40°E, 30°N-80°N) for the 108 years of ERA-20C data. Subsequently, a Poisson regression is carried out in which the track density time series per grid cell is regressed against the five defined large–scale indices. The intention for doing so is to create a "map of drivers", thus a spatial

distribution of the predominant large scale drivers accountable for windstorms in the North Atlantic domain. The predominant drivers are identified by determining the significant predictor (using a χ^2-test at a 5% significance level) with the largest absolute regression coefficient (out of the five available drivers) of the Poisson for every grid cell, respectively. Jointly the five most common drivers of the serial clustering model appear as the predominant driver in over 95% of all grid cells. Results for the grid box level analysis will be presented in Chapter 3.4.2

3.4 Results

3.4.1 Identified large scale drivers for serial clustering for 7 European regions

The large scale drivers for the seven European regions that are identified by the Poisson GLM AIC approach are presented in Table 3.4.1. The magnitude of the regression coefficient quantifies the relative importance of every selected drive. Thus, if, for example, regression coefficient β_2 for the SCA pattern increased by 1 unit, the impact on the statistical model would be $exp(\beta_2 + 1)$ times higher. As a result this implies that the larger the regression coefficient, the higher the relative importance of the associated large scale driver. The regression coefficient β_1 associated with the linear time trend is significantly positive across all seven regions, thereby confirming the identified trend in windstorm in Befort et al. (2016) for the entire domain.

Across all regions the SCA, the EA and the (station-based) NAO index appear as the main drivers of serial clustering of windstorms as well as for the drivers of active and inactive seasons. The temperature gradient between the North American continent and the Western Atlantic SSTs (introduced by Wild et al., 2015) also proves

Table 3.4.1: *Selected drivers for the Poisson GLM modelling the serial clustering of windstorms. The drivers are sorted by the magnitude of their regression coefficient. Values above the dashed line represent positive values; values below the line are negative. The observed and modelled dispersion score D is given at the bottom of the table.*

	BI (Region 1)	C.Eur (Region 2)	BI/C.Eur (Region3)	Scand (Region 4)	E.Eur (Region 5)	IP (Region 6)	Med (Region 7)
Selected large scale drivers	NAOstat SCA Tdif.Nam - - - EA POL	SCA Tdif.Nam QBO30 - - - EA EA.WR	SCA NAOstat Tdif.Nam EOF 10 - - - /	SCA W.Atl.T WestPac - - - PNA POL	SCA EOF10 W.Atl.T EA.WR - - - PNA EA.WR EA	EP.NP Tdif.Nam SCA - - - POL NAOstat	SCA EA SeaIce - - - WestPac
Dispersion score D model	1.14	0.12	0.60	1.54	0.63	0.09	-0.30
Dispersion score D observed	1.94	0.85	1.35	2.67	1.53	0.97	0.51

to be of importance related to windstorm clustering especially for the British Isles, Central Europe and the Iberian Peninsula. This is in accordance with their findings of a statistically significant correlation between this temperature gradient and windstorm counts over the Iberian Peninsula (Wild et al., 2015). The selection of the QBO in 30 hPa for Central Europe suggests a Troposphere-Stratosphere coupling that has an impact on the windstorm frequency for certain parts of Europe. Potentially there is a link between the northern hemisphere polar vortex and the Arctic Oscillation which is in turn coupled to the NAO (Baldwin and Dunkerton, 2001).

Across all regions, except for the Iberian Peninsula and the British Isles, the SCA index appears as the main driver for windstorm variability. This is somewhat intriguing as previous studies (e.g. Donat et al., 2010 or Pinto et al., 2009) identified

the NAO as the prominent driver for European storminess. The NAO is identified as the leading driver for the British Isles which could lead to the conclusion that the influence of the NAO variability pattern is not as far stretched across Europe as the SCA pattern. The anti-correlated impact of the NAO on serial clustering over the Iberian Peninsula supports the assumption of the NAO only having a significant impact for the far Western parts of Europe. To further investigate the spatial distribution of some of the large scale drivers we refer the reader to Chapter 3.4.2 where we present the result for the grid cell based analysis for some selected large scale drivers („map of drivers“).

In fact, most of the selected drivers originate from the EOF analysis of MSLP data in the northern hemisphere. An intriguing result is the selection of the northern hemispheric sea ice cover which is linked with clustering in the Mediterranean region which will be discussed in Chapter 3.4.2.Interestingly there is no negatively correlated driver identified for the joint region of the British Isles and Central Europe although there is for the individual regions. This could be due to the increased size of the region and the mixing of drivers that would only influence the Western or the Eastern parts of that region. Situations like these motivate the map of drivers presented in Chapter 3.4.2. Although focussing on Europe and the North Atlantic there are some variability modes, which have their centre of action more towards the Pacific region (i.e. PNA, West Pacific mode, EOF 10). This emphasises the importance of global teleconnections for winter windstorms, implying that in order to thoroughly comprehend the European winter climate analysis solely based on Europe might not suffice.

The dispersion index D is positive for all seven regions as given in the bottom two rows of Table 3.4.1. The largest observed values for D are found for Scandinavia, the British Isles and Eastern Europe implying that the time series of windstorm counts in these regions are following a negative binomial distribution, thus featuring larger variance than mean. The modelled values for D are too small for all regions which connote an underestimation of the variance by the statistical model. The index of dispersion is close to zero for the Iberian Peninsula and negative for the Mediterranean region denoting that the variance there is smaller than the long term mean or, in statistical terms, the process is more regular than random.

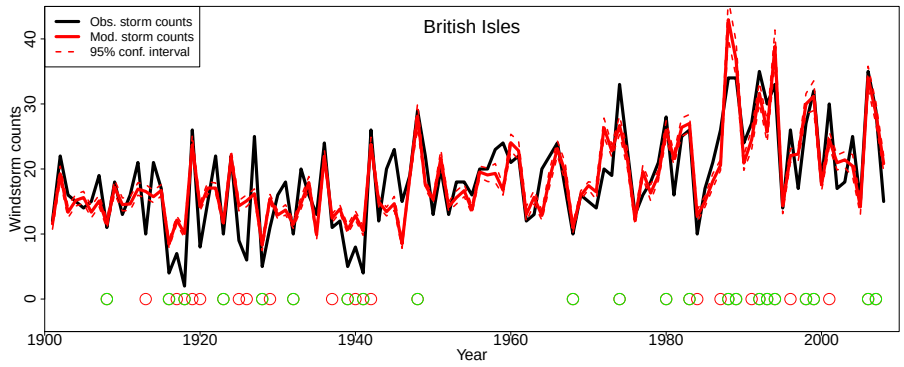


Figure 3.4.1: *Windstorm frequency per year for the British Isles from 1901–2008. The observed counts are depicted in black; the Poisson GLM modelled counts in red. The red dashed line represents the 95% confidence interval. Circles indicate active/inactive seasons. Green circles mark seasons in which the active/inactive seasons are predicted correctly whereas red circles indicate years where the activeness of the season was not predicted correctly.*

The qualitative performance of the model can be assessed in Figure 3.4.1. It depicts the modelled (cross–validated) as well as the observed windstorm counts per winter season for the British Isles region (Region 1). The observed time series (black) is within the 95% confidence intervals for most of the years implying a good fit of the model. The large amount of green circles further implies the capability of the model to skilfully identify active and inactive seasons.

The quantitative performance of the cross validated statistical model for the seven regions is displayed in Table 3.4.2. The average correlation between modelled and observed time series over all seven regions of 0.75 is remarkably good considering the application of a relatively simple Poisson GLM. The model works particularly well for the British Isles and the Scandinavian region where the correlation between modelled (cross–validated) and observed time series is larger than 0.80. Furthermore the model explains up to 70% of the variability of the inter-annual storminess.

The British Isles and Scandinavia are also among the regions featuring the highest active and inactive season hit rates. A stellar example is the 79% active season hit rate for the British Isles. The best inactive season hit rate is found for the Mediterranean area. As discussed before, the modelled dispersion is generally smaller than the observed value throughout which is due to the reduced variability in the model. Although assuming a Poisson process for the windstorm counts, the dispersion statistic is unequal to zero. If it was zero, no clustering would be observed. Thus this deviation from zero is exactly what shows that most of the time series feature clustering.

3.4.2 Map of drivers

Taking literature and our results from 4.1 into account we chose the NAO, the SCA, the EA, the POL and the sea ice time series as the overall leading variability patterns associated with winter windstorms for the European, the North Atlantic and the Mediterranean domain. The "map of drivers" created by grid cell based Poisson GLMs is presented in Figure 3.4.2a. The most important drivers for the variability

Table 3.4.2: *Results for the cross-validated metrics of the Poisson model performance. The highest value in each column is marked bold ED stands for explained deviance, an equivalent to the R^2 value for linear regression models.*

Region	*Hit rate active*	*Hit rate inactive*	*Disp. Score D bias*	*Correlation/ED*
British Isles	**78.9%**	61.1%	-0.80	**0.84/0.70**
Central Europe	71.1%	50.0%	**-0.73**	0.72/0.51
BI+C.Europe	71.1%	65.5%	-0.75	0.81/0.64
Scandinavia	75.0%	64.3%	-1.13	0.82/0.67
Eastern Europe	68.8%	43.8%	-0.89	0.76/0.57
Iberian Pen	56.3%	41.2%	-0.88	0.66/0.44
Mediterranean	22.2%	**77.6%**	-0.81	0.63/0.40
Mean	***63.3%***	***57.6%***	***-0.86***	***0.75/0.56***

of winter windstorms over main parts of the European mainland are the SCA and the NAO time series. The British Isles appear to be under the influence of several predominant predictors. Whereas the southern part of the British Isles is influenced mostly by the SCA pattern, the northern part is more affected by the NAO. This is in line with the findings of Chapter 3.4.1 where we showed that the leading driver for serial clustering over the British Isles is the (station-based) NAO index time series.

Due to co-linearity naturally only one of the two NAO time series can be present in the pool of potential drivers. Although the EOF based NAO exhibits a correlation of over 0.8 with the station-based index, the station–based NAO index yielded the better results with regard to the statistical model. The right panel of Figure 3.4.2 depicts the Explained Deviance (ED; McCullagh and Nelder, 1989) function of the Poisson distribution which represents an equivalent to the R2 value of a linear regression, i.e. it estimates how much variance is explained by the Poisson model in every grid cell. It is particularly high over the Northern Sea, Scandinavia and the East Atlantic with up to 60% variability jointly explained by the five large scale drivers. Generally, more windstorm variance is explained over the northern parts of

Europe compared to the southern parts. The extent of the northern hemispheric sea

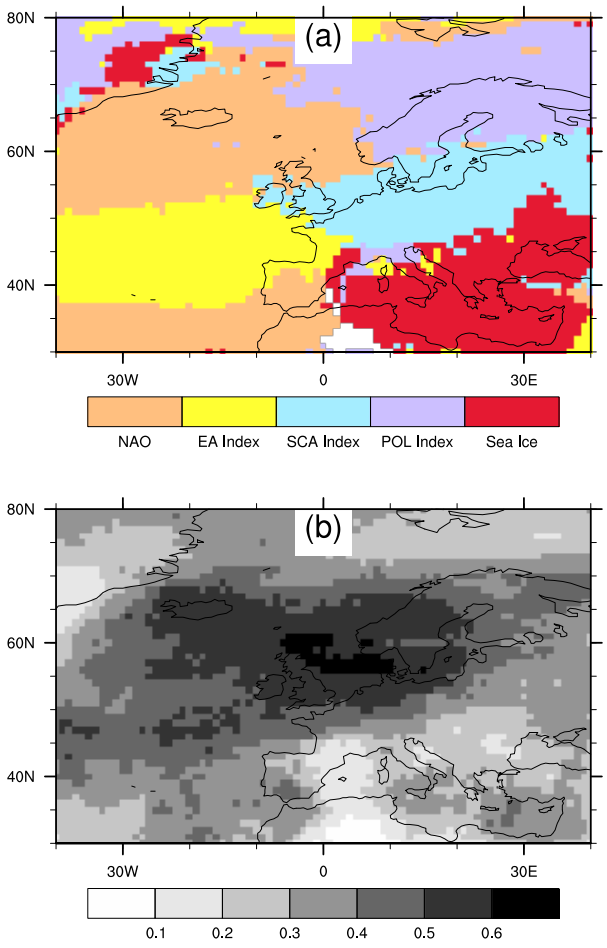


Figure 3.4.2: *(a) Most dominant teleconnection patterns explaining the inter-annual variability of windstorm counts per grid cell from 1901-2008 identified by a Poisson GLM. (b) Explained deviance of the Poisson regression model for every grid cell*

ice cover operates as a significant driver of storminess particularly for the Central

to Eastern Mediterranean area and parts of the Black Sea. This is in line with the findings of Chapter 3.4.1 where sea ice appeared as the second most important driver in this region, thereby also justifying the selection of sea ice as one of the main drivers for this grid cell based model.

An explanation for the sea ice as a significant driver could be the enhanced (weakened) meridional temperature gradient with more (less) sea ice extent as it is one of the factors that controls the stability of the mid-latitude flow. Semmler et al. (2016) found a decreased number of wintertime cyclones in their sensitivity analysis with an 80% reduced Arctic sea ice extent over the Eastern Mediterranean and parts of the Black Sea and also a reduced Eady growth rate over the Balkan region and the Black Sea (their Figures 12f and 14f respectively). Our results indicate that larger sea ice extent leads to more windstorms, which is generally in line with those findings from Semmler et al. (2016). However, the interpretation of sea ice as a factor steering windstorm activity over this region should be done with caution as only about 10-20% of the windstorm variability is explained by sea ice variability (Figure 3.4.2 right panel). It's also worth mentioning that there are several studies showing an opposite relationship between sea ice and cyclone/storm activity over the Mediterranean region (Grassi et al., 2013). This indicates that further targeted studies are necessary to fully understand the link between northern hemisphere sea ice extend and storm activity over southern Europe, which is however beyond the scope of this study.

As its name implies, the Polar index (POL) serves as the main influence on inter-annual windstorm counts in northern Scandinavia and the Polar regions. The EA

index dominates the region between the two NAO poles which makes it the dominant driver for the Eastern Atlantic as well as parts of Western Europe, explaining the major part of the 60% inter-annual variability for that area. The boundary between the NAO and the EA index, especially in the south, is remarkably well defined, implying that the area of influence of either of the predictors can be localised very precisely. It has to be noted that the left panel of Figure 3.4.1 only shows the "winning" large-scale index yielding the largest absolute regression coefficient, for every grid cell. Most of the grid cells in the North Atlantic domain, however, feature more than one significant predictor. As the absolute value of the coefficients is depicted in Figure 3.4.2, it also has to be considered that some of the teleconnections (regression coefficients) are positive (northern pole of the NAO, SCA) and some are negative (southern pole of the NAO, EA index; cf. Figure 3.4.3). The panels in Figure 3.4.3 depict the regression coefficients of the five chosen teleconnection patterns and for the linear time trend as calculated by the Poisson GLM. The NAO exhibits the prominent dipole structure associated with the Icelandic Low and the Azores high. Its area of significant influence stretches across most parts of the Atlantic, the Iberian Peninsula, the British Isles and some parts of western Scandinavia. Strikingly, the European mainland is almost solely under the influence of the SCA teleconnection pattern (Figure 3.4.3c). This can be better understood by comparing the patterns of the two variability modes (Figure A.2.1 in the Appendix). The SCA and the NAO patterns look somewhat similar consisting of a negative pole in the north and a positive pole in the south. The SCA index, however, is shifted downstream so that its main centres of action are over Scandinavia and the Iberian Peninsula. This shifted dipole enables storm tracks to travel more directly eastwards and stretch further into the European mainland whereas the NAO leads storms on

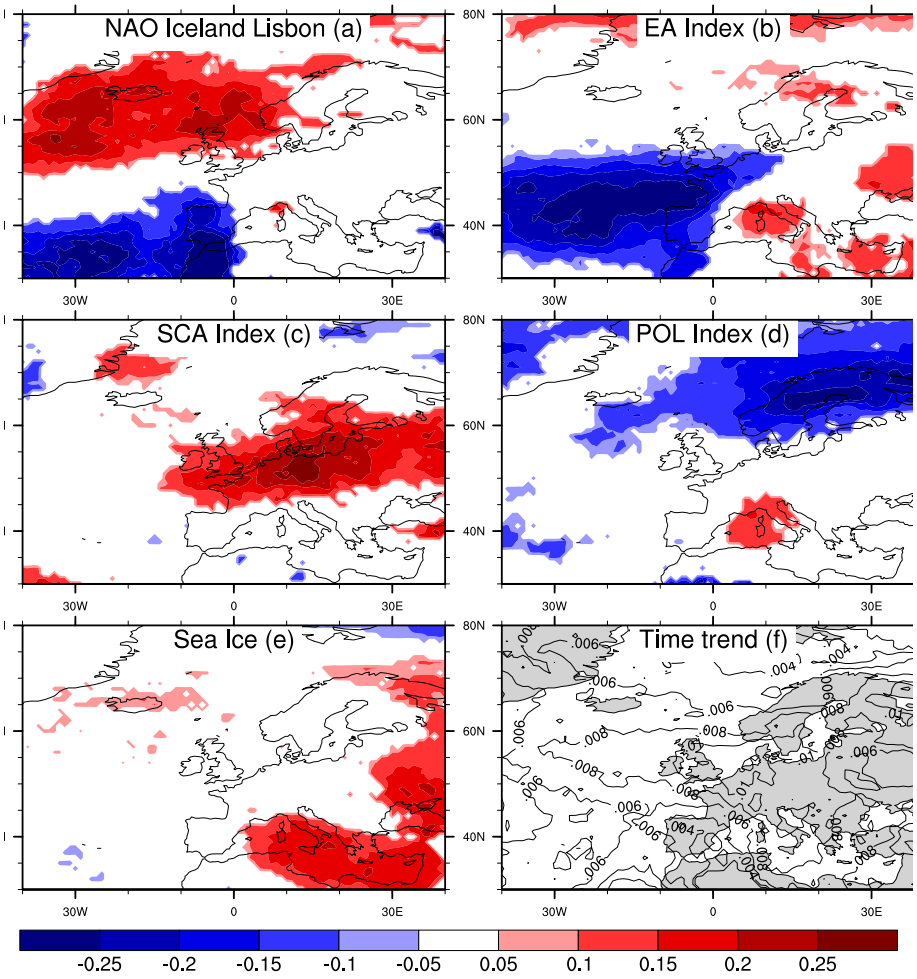


Figure 3.4.3: *Significant Poisson GLM regression coefficients for the dominant selected large scale drivers explaining most of the variability of winter windstorms for Europe/North Atlantic (a)-(e) Time trend coefficients of the Poisson GLM as contours (f). The contour levels for the time trend are separated by 0.002 intervals.*

a more north easterly trajectory across Scandinavia.

The regression coefficients of the POL index also appear as a weak dipole with its centres of action over the Mediterranean Sea and Scandinavia up to the Polar Regions. The EA index is the dominant negative teleconnection pattern between the two dipoles of the NAO. Thus, during its negative phase the likelihood of observing more windstorms in the East Atlantic and Western Europe is significantly increased. The area of significant coefficients coincides well with the positive centre of the EA pattern (compare Figure A.2.1 in the Appendix). The time trend is positive across almost the entire domain which implies a positive trend in windstorm counts. The strongest trend is found over the British Isles and Scandinavia whereas the trend over the Mediterranean area is considerably smaller.

In order to check for non-stationary teleconnections, especially with regard to the Pacific (e.g. Greatbatch et al., 2004), the drivers were also examined for two halves of the century separately (1901-1950 and 1951–2008; not shown). The results look very similar to the map presented in Figure 3.4.2a so that we can assume stationary teleconnection processes.

3.5 Summary and Discussion

This study identifies and quantifies the impact of various large scale drivers on European winter windstorm counts and their seasonal clustering. For this purpose windstorm events were identified using an objective wind tracking algorithm (Kruschke, 2015 and Befort et al., 2016) in the 20th century ECMWF reanalysis ERA-20C (Poli et al., 2016). Befort et al. (2016) have shown that the high frequency variability of windstorms in ERA-20C and NOAA 20CR (Compo et al., 2011) is well correlated so that we can assume that the windstorm climatology in ERA-20C and NOAA

20CR are in sufficient agreement for the scope of our study. We are fully aware of the potential issues regarding the century-long reanalysis efforts especially for the beginning of the century. However, with this study we are not trying to validate and assess the quality of the reanalysis but solely trying to understand the drivers for the inter-annual variability of windstorms.

The windstorm trajectories were used to count the annual amount of windstorms for seven defined European regions representing i) diverse windstorm climatologies and ii) regions of interest within the actuarial community. The statistical model developed in Chapter 3.3 is aimed at modelling serial clustering of windstorms over seven different European regions. The model is capable of assessing the inter-annual variability in two different perspectives:

(i) The clustering dispersion: Thus whether or not the variance-to-mean ratio is larger than 1 (according to Mailier et al., 2006). In this case the time series is over dispersed which in return corresponds to the fact that occurrences (in this case storm counts) occur in clusters. The benefit of this approach is the understanding of which physical drivers/large scale modes are directly responsible for the deviation from the Poisson distribution.

(ii) The modelling of the active/inactive seasons: Whether or not a single season spawns more or less than one standard deviation of storms compared to the long term mean. Ultimately, a skilful model w.r.t. modelling active/inactive seasons could be particularly beneficial for the insurance sector as windstorms over Europe are associated with extensive losses, especially if they occur in quick succession which is more likely in seasons with high storm counts.

The model is developed using a stepwise Poisson AIC approach to select the domi-

nant drivers out of a pool of 20 large scale drivers that entail MSLP and SST related time series (cf. Table 3.2.1). Compared to previous clustering studies (e.g. Mailier et al., 2006); the investigated time series contained the entire 20th century, so that the results comprise a longer time scale than before. The regression coefficient associated with a linear time trend is significantly positive across all seven regions, hereby confirming the results of Befort et al. (2016).

Generally the statistical model shows very satisfactory results regarding the goodness of fit of the windstorm count, the estimation of the dispersion score of the time series, and the ability to successfully define a respective season as active or inactive across all seven regions. Especially the hit rates of predicting active seasons for most of the regions are very promising. The results of this study suggest, that a determinant proportion of the information, needed to accurately describe windstorm frequency and clustering, is being efficiently extracted from the set of large-scale indices discussed, especially placing value on those indices selected for the different regions. The calculated scores confirm that the model is able to perform well even in regions with strongly diverse windstorm climatologies, e.g. in the Mediterranean area or over the Iberian Peninsula. Although a lower overall performance is observed for those regions, most of the metrics assessed still show a satisfactory result (correlation >65% for the Iberian Peninsula), and more skill than a random forecast of hit rates (>50% for the Mediterranean).

The five predominant drivers identified in literature and with our regional analysis were investigated in more detail on a grid box level via a Poisson GLM approach in order to produce a "map of drivers". This map of drivers represents a spatial

distribution of the predominant large scale driver for every grid cell across Europe and the North Atlantic. Thus, it can be used to deduct the influence of a respective driver for winter windstorm variability. These five predominant drivers entail NAO, SCA, POL, NH sea ice cover and the EA time series. These five teleconnection patterns represent significant drivers for more than 95% of all grid cells and jointly explain up to 60% of the inter-annual windstorm variability (Explained Deviance) over large parts of Northern Europe and the Northern Sea.

This result is in good accordance with Seierstad et al. (2007) who also identify four out of these five large scale teleconnections with regard to European winter storminess. Instead of using actual tracked windstorm events, however, their storminess is defined as high-pass filtered MSLP variance. Hence we can confirm their findings based on a wind speed based definition of storminess. Their study reveals that the EA/WR pattern also has a significant impact on European storminess. Even though the EA/WR pattern is not found to be having an essential impact on grid cell level it is identified as being significant for the statistical model for 2 out of the 7 European regions which is in turn in overall accordance to their findings.

The identified drivers both for the impact and the physical based analysis are in good agreement with previous windstorm and windstorm clustering literature, confirming the SCA, the NAO, the EA and the POL time series as major drivers of the variability (c.f. Mailier et al., 2006, Seierstad et al., 2007, Vitolo et al., 2009, Pinto et al., 2009, Donat et al., 2010). A striking result is represented by the importance of the SCA pattern both in the map of drivers and the impact analysis as it is identified as the most important driver for Central Europe. This implies that when assessing

the winter windstorm hazard in (Central) Europe it seems as if it is not sufficient to focus on the NAO alone. The EA pattern is, as its name implies, the dominant feature for the East Atlantic including the northern part of the Iberian Peninsula and the southern part of the British Isles. The teleconnections can assumed to be stationary as the results for the split century analysis (1901-1950 and 1951-2008) show very similar results to the map of drivers comprising the entire century from 1901-2008.

Additionally to the previously studied drivers, we were able to identify the temperature gradient between the North American continent and the West Atlantic (Wild et al., 2015) and the northern hemispheric sea ice extent as locally important drivers (Iberian Peninsula, Mediterranean). Sea ice cover as a teleconnection with the Mediterranean is generally in line with results from Semmler et al. (2016) who found a reduced number of cyclones in the Eastern part of the Mediterranean and over the Black Sea in their sensitivity study with reduced sea ice. We argue that we observe the same direction of that correlation, thus more sea ice could lead to an increased Eady growth rate and higher numbers of windstorms in the Eastern Mediterranean and the Black Sea area. We assume that the storminess in this area is linked with an enhanced meridional temperature gradient over the North Atlantic. Sea ice, however still has to be treated with caution as we could show that it only makes up for about 10-20% of the explained deviance in the respective region. The maximum of this explained deviance can be found in areas where there is also some influence from the EA and the POL which both show weak dipole structures over the Mediterranean (cf. Figure 3.4.2b and 3.4.3b, d).

It should be noted that neither the model developed on grid box level nor the model for the seven regions include interactions between the predictors. An inclusion of those might improve the fit/skill of the regression/model; however, it proves to be more demanding to identify the impact of individual predictors. For that reason, the interactions between the predictors have not been considered. However, regarding future work this is definitely an approach worth investigating in more detail, especially for the development of a more sophisticated statistical model. Considering the time trend of the Poisson model, a quadratic or even exponential trend could be tested and compared to the linear trend.

Albeit the model being of relatively simple nature, it produces more than satisfactory results, particularly with regard to the identification of large scale drivers for the seven defined regions and the hit rates for most of the seven regions (correlations $>80\%$, HR$>70\%$ for important North Atlantic regions). Arguably there is a large overlap of predictors between some of the regions; however, none of the regions feature the exact same predictors. Some regions, e.g. the Mediterranean box, could perhaps be subdivided in a western and eastern part, as the grid box level regression suggested that sea ice for example only impacts the eastern part of this box. In that way the predictors could be determined even more precisely.

The added value of the presented paper is given by revealing a more comprehensive insight into the physical drivers for serial clustering of windstorms over the European continent on a time scale of more than 100 years. High quality wind field forecasts are rarely freely available, whereas large scale indices (i.e. NAO) are often freely available for the public to download. This is why the work and in particular

the model presented here could prove to be beneficial for the insurance sector as results suggest that fairly simple Poisson GLMs are already able to skilfully estimate whether or not a seasons has the potential of spawning an increased or decreased amount of winter windstorms. Thus, by using large-scale indices, a potential degree of clustering and the number of expected windstorms could be directly obtained from such a model for the subsequent winter season. Additionally the map of drivers that was presented in Chapter 3.4.2 provides a useful overview over the spatial structure of prominent large scale drivers for the European domain.

Chapter 4

Predictability of extreme wind speeds associated with windstorms

"Too bad the post office isn't as efficient as the weather service"

– Dr. Emmet Brown, crazy scientist from Back to the Future II (1989) in the fictional future year 2015

Abstract

As extreme wind speeds are responsible for large socio-economic losses in the European domain, a skillful prediction would be of great benefit for disaster prevention as well as the actuarial community. Here we evaluate the patterns of atmospheric variability and the seasonal predictability of extreme wind speeds (e.g. $>95^{\text{th}}$ percentile) in the European domain in the dynamical seasonal forecast system ECMWF System 4 and compare to the predictability using a statistical prediction model. Further we compare the seasonal forecast system with ERA-Interim in order to advance the understanding of the large-scale conditions that generate extreme winds. The dominant mean sea level pressure (MSLP) patterns of atmospheric variability show distinct differences between reanalysis and System 4 as most patterns in System 4 are extended downstream in comparison to ERA-Interim. This dissimilar manifestation of the patterns across the two models leads to substantially different drivers associated with the generation of extreme winds: While the prominent pattern of the North Atlantic Oscillation (NAO) could be identified as the main driver in the reanalysis, extreme winds in System 4 appear to be related to different large-scale atmospheric pressure patterns. Thus, our results suggest that System 4 does not seem to capture the potential predictability of extreme winds that exists in the real world. This circumstance is likely related to the unrealistic representation of the atmospheric patterns driving these extreme winds. Hence our study points to potential improvements of dynamical prediction skill by improving the simulation of large-scale atmospheric variability.

The following chapter is an edited version of the previously published article:
Walz, M. A., Donat, M. G., & Leckebusch, G. C. (2018). Large-scale drivers and seasonal predictability of extreme wind speeds over the North Atlantic and Europe. Journal of Geophysical Research: Atmospheres, 123. https://doi.org/10.1029/2017JD027958

The idea for this study was developed by Michael Walz, in particular the use of the statistical entropy. The outline of the study was designed by Michael Walz and Markus Donat. All data processing and analysis was carried out by Michael Walz. The manuscript was written by Michael Walz with Markus Donat and Gregor Leckebusch contributing to the writing.

4.1 Introduction and Motivation

Winter windstorms represent one of the most dangerous and loss-intensive natural hazards for the European region. According to the European Environmental Agency (EEA), storms were the costliest natural hazards in Europe between the years 1998 and 2009 exhibiting an accumulated loss of more than €44 billion (EEA, 2011). Thus, it would be of utmost value to provide useful predictions on seasonal scales as it would enable decision makers to take measures in order to minimize potential losses and most importantly to avoid casualties.

The demand for these longer term "weather forecasts" exceeding the common 10 day prediction period has generally increased considerably over the last decade. One of the reasons for that is certainly the desire to minimize casualties and loss due to extreme weather events especially with respect to climate change (e.g. Easterling et al., 2000; Lowe et al., 2016). Due to the atmosphere's chaotic nature however, it is generally impossible to predict single weather events, not to mention extreme events, deterministically on a time scale exceeding 5-7 days. The reason for that is the non-linearity of the atmospheric system which amplifies minuscule deviations in initial conditions into large disturbances at the end of a forecast period (Lorenz, 1963). This behavior of the atmosphere is often referred to as the "Butterfly effect". There is, however, an intrinsic predictability in atmospheric variables on a longer time scale. This predictability is dependent on atmospheric and oceanic conditions that feature variability modes on longer time scales (e.g. the El Niño-Southern Oscillation (ENSO or the Atlantic Multidecadal Oscillation (AMO), e.g. Knight et al., 2006). The AMO is known to have an impact on the decadal variability of the North Atlantic storm track (Nissen et al., 2014) and also on its position (Woollings and

Blackburn, 2012).

As a response to the rising demand of forecasts for intermediate time scales, the gap between day-to-day weather forecasts and long-term climate projections has been closed by seasonal predictions (Palmer et al., 2004; Doblas-Reyes et al., 2013). These seasonal predictions are based on free running coupled atmosphere-ocean models that are usually initialized on the first day of each month. The initial conditions are taken from the observed state of the atmosphere and the ocean on the given day. In most cases seasonal forecasts run for a period of 6-7 months. To account for the chaotic nature of the atmosphere they are run as ensemble forecasts which often feature 20-50 ensemble members to enable probabilistic predictions.

As for every forecast product, there is always the question of forecast skill. Della-Marta et al. (2010) investigated the skill of the ECMWF System 3 (Anderson et al., 2007) with regards to high wind speeds and found skill for the 95^{th} percentile of wind speeds for the first lead month but none thereafter.

Torralba et al. (2017) examined the skill of ECMWF System 4 (Molteni et al., 2011; System 4 hereinafter) regarding the potential wind energy yield (thus seasonal mean wind conditions) for two different regions featuring a substantial amount of wind farms (Canada and the North Sea). They showed that the seasonal mean wind predictions of System 4 have to be bias corrected in order to be useable for the end user (i.e. wind farms). The bias corrected seasonal mean wind speeds, however show skill especially in the tropics but also in areas relevant to wind energy production, e.g. extra-tropics. MacLeod et al. (2018) also investigated the wind energy yield by

performing statistical simulations based on sub-daily (6-hourly) to monthly mean wind speeds taken from two reanalysis data sets. They concluded that only if there is predictable information in daily values there is an information gain in using (sub-) daily values compared to weekly to monthly means. As the predictability in seasonal forecast products is based on low-frequency variations of the climatic system they are expected to have more skill on a weekly to monthly scale rather than a daily scale.

Extreme and mean winds in Europe are related to the North Atlantic Oscillation (NAO; e.g. Hurrell, 1995) as the leading mode of atmospheric variability for Europe (e.g. Pinto et al., 2009; Donat et al., 2010) especially for the winter months. Thus a skillful prediction of the NAO could provide valuable information on potential storm related impacts for upcoming winter seasons way ahead of common weather forecasts.

The UK Met Office Global Seasonal forecast System 5 (GloSea5; MacLachlan et al., 2015) has been shown to have significant skill for predicting the NAO for the European winter period (Scaife et al., 2014). Eade et al. (2014), however, argued that GloSea5 (and potentially other systems) underestimates the inherent predictability of the climate system (including the NAO) and that single ensemble members contain too much noise. They proposed the use of large ensembles in order to reduce noise and also a post processing method to adjust the variance of the ensemble prediction. Recently O'Reilly et al. (2017) investigated century long seasonal hindcast simulations (uncoupled with prescribed SSTs and System 4 based, see Weisheimer et al., 2017) and found a high variability in prediction skill of the NAO and the Pacific/North American Index (PNA). In particular they identified a strong link be-

tween the lack of skill for the PNA and the SST anomalies in the Pacific during the mid-twentieth century in the seasonal hindcast experiments (for this particular case).

Moving from the potential predictors of winter windstorms towards the actual event, Befort et al. (2018) investigated the capability of GloSea5 and System 4 to predict the frequency of occurrence of winter windstorms per season. They found significant positive correlations between the observed and forecasted number of windstorms on a grid cell level in large areas over the Atlantic Ocean for the years 1992-2011. However, they observed a drop in significance when investigating the skill for the longer 1983-2011 time span.

Walz et al., 2018a in turn showed that a considerable amount of the inter-annual variability of these winter (DJF) windstorms counts can be explained by large scale Atlantic drivers (e.g. NAO, Scandinavian Pattern; SCA or the East Atlantic Pattern; EA). Our study can be considered as a nexus between these two before mentioned papers as a direct link between large scale variability patterns and European winter extreme winds (in seasonal forecasts in particular) has not been shown to date. In order to investigate the predictability of extreme wind speeds we are therefore trying to answer the question in how far the NAO (or other large scale variability patterns) are associated with extreme wind speeds (associated with winter windstorms) in System 4. Thus, there are four questions we are trying to answer in this study:

- How well are winter large-scale MSLP (e.g. NAO, SCA) patterns generally represented in the ECMWF System 4?
- What is the predictability of extreme wind speeds in seasonal forecasts? In other words, does the forecast contain more information than the climatology?

- Which large scale-drivers are responsible for extreme wind speeds in System 4 and in the reanalysis respectively?
- How does a statistical model compare to the dynamical seasonal forecast regarding the predictability of extreme wind speeds to occur?

The next section of this paper features a description of the seasonal forecast system and reanalysis product used for this study. Chapter 4.3 introduces the methods used for this analysis. This includes the Empirical Orthogonal Function (EOF) analysis of the MSLP data, the concept of statistical entropy as a skill measure, the quantification of the extreme wind speeds and the description of the statistical model used to predict extreme wind speeds for the December to February (DJF) period employing MSLP data from October/November. The results of this analysis will be presented in Chapter 4.4 before concluding this paper with a summary and discussion of these results.

4.2 Data

The predictability of extreme wind speeds (using 6-hourly wind speed) and MSLP variability modes in the retrospective forecasts is investigated for the ECMWF Seasonal Prediction System 4 (System 4; Molteni et al., 2011). 6-hourly wind speeds have been used extensively to assess extreme winds for example associated with winter windstorms (e.g. Nissen et al., 2014 or Leckebusch et al., 2008a). As the focus is on wind speeds associated with winter windstorms boreal winters from 1982/83 until 2013/14 are used which results in 32 DJF winter seasons.

The reforecasts are initialized every 1st of November which corresponds to a lead

time of 1-4 months when analyzing DJF. The atmospheric initial conditions are taken from the ERA-Interim reanalysis (Dee et al., 2011). The ocean is initialized using the Ocean ReAnalysis System 4 (ORAS4; Molteni et al., 2011). Every re-forecast comprises 51 ensemble members and is provided on a spectral resolution of T255 which is the same resolution used for ERA-Interim. The perturbed initial conditions are produced using singular vectors and an ensemble of ocean conditions of the ocean model NEMO (Madec, 2008). The forecast system is based on the IFS cycle 36r4 of the ECMWF.

In the absence of high-quality, homogeneous wind observations for a number of locations across Europe, the predictability/skill of System 4 is assessed against ERA-Interim. We note that, as both System 4 and ERA-Interim are based on similar model versions at the same resolution, there is only a minor bias in the simulated wind speeds, which should favor the detection of prediction skill. The wind fields for the North Atlantic/European domain (70°W-40°E, 30°N-70°N) are remapped to a 5x5 degree grid for the predictability assessment to minimize the effects of small-scale spatial noise, in line with recommendations for decadal prediction verification (Goddard et al., 2013).

4.3 Methods

4.3.1 EOF Analysis of MSLP data

EOFs are described by the eigenvectors of the covariance matrix of time series at different spatial points (e.g. Jolliffe, 1986; Ambaum et al., 2001). By definition all eigenvectors are orthogonal to each other. EOFs represent an ideal tool to investi-

gate the spatial and temporal variability of any given variable as each EOF explains the maximum temporal variance among all spatial fields. Usually the EOFs are ranked by the amount of total variability that they explain. When using Northern Hemispheric (NH) MSLP data, the first EOF or eigenvector usually corresponds to the NAO as it is responsible for most of the variability in the NH (e.g. Ulbrich and Christoph, 1999 or Woollings et al., 2015). The second mode is usually associated with the Pacific North American pattern (PNA, e.g, Ambaum et al., 2001; naming of the EOF modes in the reanalysis in accordance with NOAA, `http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml`) which is the leading pattern of inter-annual variability of MSLP for the Pacific region.

For this study the EOF analysis was carried out for the standardized (by dividing by the standard deviation) DJF 6–hourly MSLP anomalies of the entire NH (20°N-90°N) for the ECMWF System 4 and ERA-Interim to compare their spatial patterns. Thus, in the case of System 4 for example, all 6–hourly MSLP values of all 51 ensemble members are thrown together to obtain the 9 leading EOFs. For ERA–Interim the EOF is simply computed from all 6–hourly MSLP data over all DJF seasons.

As it turns out, the patterns of the two models look very different (see Chapter 4.4.1 for details). This could be due to the different dynamics within the reanalysis and the seasonal forecast respectively. In order to obtain comparable EOF time series for further analysis (especially with regards to drivers of extreme winds) we projected the MSLP anomalies of System 4 onto the eigenvector loadings of ERA-Interim. The projection is a spatial projection onto the eigenvector patterns. That way we

are aiming for a fairer comparison between the different driver time series of Season 4 and ERA-Interim respectively. For completeness the results employing the System 4 eigenvectors are shown in the Appendix.

For the process of developing the statistical prediction model the EOF analysis was also done for the same region for MSLP data from October 15th to November 15th for every year for ERA-Interim. In order to make use of the large ensemble and to reduce the noise of each individual ensemble member, all 51 members are pooled together to calculate joint EOFs. If the EOF was calculated for each of the members individually, the variability of the ensemble mean would be too low (Figure 4.4.8). The first nine EOFs were calculated for both System 4 and ERA-Interim. The scores or principal components (PCs) of the EOF analysis represent the time series of the respective EOF pattern. For comparison reasons the NAO index was also calculated as a normalized pressure anomaly difference between the grid boxes closest to Lisbon and Reykjavik respectively from the ensemble mean of the hindcasts in System 4.

4.3.2 A normalized sum as a measure for extreme wind speeds

As extreme wind speeds (represented by percentiles) tend to be of fairly erratic nature with large year-to-year variability, the percentile alone does not show a high predictability on an inter–annual time scale (see Figure A.3.1 in the Appendix). Extreme wind speeds here refer to seasonal exceedances of percentiles calculated from the climatological 6–hourly wind speed distribution. For that reason we chose to estimate the predictability of System 4 extreme wind speeds by incorporating the entire upper tail of each local wind speed distribution in order to get a more robust target variable. This is implemented by summing up all wind speed values exceeding

a certain percentile threshold and normalizing this sum with the total amount of available time steps per season n (see Equation 4.3.1). The resulting value $I_{k,y,9X}$ represents an integrated measure over the amount of percentile exceedances as well as the magnitude of exceedances on an inter-annual time scale. Accordingly a large value could either represent a season with many smaller exceedances or a season with few large exceedances of the respective percentile.

$$I_{k,y,9X} = \frac{\sum_1^n v_{n,k}}{n}, \forall v_{n,k} > v_{9X,k} \tag{4.3.1}$$

Here the $X = 0, 5, 8$ and not to be confused with the X in Equation 4.3.2. We chose the 90^{th}, the 95^{th} and the 98^{th} percentile as thresholds for this study as these percentiles have been used previously with regards to extreme wind speeds (e.g. Leckebusch et al., 2008a or Della-Marta et al., 2010). The percentile is calculated as a climatological percentile from the mutually available data (1982-2014) for forecasts and observations separately in order to avoid reducing predictability by a potential bias. $I_{k,y,9X}$ is calculated for every grid cell k for every DJF season y between 1982/83 and 2013/14 for all 51 ensemble members and compared to ERA-Interim for all three percentiles. In order to assess the predictability in the European region, $I_{k,y,9X}$ is averaged for time series comparison for the box shown in Figure 4.4.6.

4.3.3 Statistical Entropy and Predictive Power

The concept of statistical entropy has its roots in information theory and was first introduced by Shannon (1948, 1951) who defined it as a measure to estimate the information content of a message. Schneider and Griffies (1999) and DelSole (2004) took up on this idea and developed a conceptual framework of how to use statistical

entropy as a measure of predictability. They argue that a forecast is best described by its entire distribution and that valuable information is lost when the distribution is reduced to mere moments (e.g. RMSE).

As this section is merely adequate to do justice to the complex field of predictability and information theory we refer the reader to the excellent original papers by Schneider and Griffies (1999), DelSole (2004) and also Tang et al. (2007) for a detailed disquisition of the topic. This section shall provide a short and more qualitative overview over the concepts that will be used in this paper. The Shannon (statistical) entropy S_X of a random variable X is defined as (Shannon, 1948):

$$S_X = -\int dx\, p_x(x) \log(p_x(x)) \tag{4.3.2}$$

S_X is a measure of uncertainty related to the random variable X and $p_x(x)$ represents the occurrence probability. If all realizations of the random variable X are equally likely, S_X is maximal as the outcome of the realization is hardest to predict. An example would be the rolling of a fair dice as every number has a probability of 1/6. Thus every roll of the dice conveys the maximum amount of information. If the dice was biased, say the number 3 was on four of the six sides, there would be less uncertainty as the number 3 can be expected to appear with a probability of 2/3. This reduction in uncertainty also reduces S_X. The extreme case scenario would be a dice with the same number on every side. The uncertainty would be zero and so would be the Shannon entropy as the rolling of the dice does not convey any new information. The Shannon entropy can also be defined for multivariate state variables by the Joint entropy.

This concept can be applied to ensemble forecasts: The uncertainty within an ensemble forecast depends on the spread of the ensemble members, and would be maximal if every member would predict a completely different state of the atmosphere. The probability of every predicted atmospheric state would be 1/51 in case for System 4. If all ensemble members of the forecast ended up in the exact same state however, there would be no uncertainty at all in the forecast and thus Shannon's entropy would be zero.

As follows, a robust forecast will produce smaller values of the statistical entropy as it is more constrained. However the actual (correlation-) skill of the forecast could still be small in the case that all ensemble members drift to an incorrect state. This would be considered an overconfident forecast. A measure of predictability that makes use of these considerations is given by the Predictive Information (PI) as defined by Schneider and Griffies (1999) and adapted by Tang et al. (2007):

$$PI = S_X - S_E \tag{4.3.3}$$

S_X describes the uncertainty (statistical entropy) associated with the climatological distribution, thus if no forecast was available (prior uncertainty). The variable SE represents the statistical entropy after the forecast (and observation) has become available. It quantifies the uncertainty in the ensemble prediction (posterior uncertainty). Ideally the posterior uncertainty is smaller than the prior uncertainty ($S_E < S_X$) due to useful information in the ensemble forecast. Hence a large PI value represents forecast which includes more information than the climatology, thus an information gain.

Practically however, the uncertainty in ensembles can be larger (i.e. $PI < 0$) than the uncertainty in the climatology especially for an erratic variable such as extreme wind. Grid cells with negative values are considered to feature no predictability/additional information to the climatology. In other words the climatology provides more information (is "sharper" than the forecast. Another measure defined by Schneider and Griffies (1999) and adapted by Tang et al. (2007) is represented by the Predictive Power (PP):

$$PP = 1 - exp(-PI) \tag{4.3.4}$$

Equivalently to the interpretation of the PI, large values of the PP suggest useful information in the seasonal ensemble forecast making it superior to the climatology. Evidently PP becomes maximal ($PP = 1$) if the prior uncertainty is infinite and the posterior uncertainty vanishes. As the PP exhibits a proper limiting behavior ($0 < PP < 1$) it represents a descriptive quantification of the predictability. For this paper the PP will be calculated for the previously defined variable $I_{k,y,9X}$ for each winter individually and also as a time mean which refers to the average predictability of extreme wind speeds for every 5x5 grid cell in the studied domain.

4.3.4 AIC selection of large scale drivers

In order to understand what drives the inter-annual variability of extreme winds in System 4 as well as in ERA-Interim we examined which of the 9 large-scale MSLP variability patterns (EOFs) are most significant for extreme winds in the North Atlantic and European region. The selection criterion is based on the AIC (An Information Theoretic Criterion, occasionally also referred to as Akaike Information Criterion; Akaike, 1974) step-wise approach. Similar to the statistical entropy, the

AIC is based on information theory (Jaynes, 1957). The AIC is an estimate of how much information is lost by using a statistical model instead of the actual physical relation. Thus, the AIC can be used as a tool for model selection if different models can be compared to each other. The essential part of the selection process is the tradeoff between the goodness of fit and the complexity of the model, as the number of parameters to be estimated (k) as well as the maximum of the likelihood function L are part of the AIC score:

$$AIC = 2k - 2\ln L \tag{4.3.5}$$

The winning model will exhibit the smallest AIC value. Different from other model selection criteria, e.g. the F-test, the AIC score does not provide any information about the absolute quality of the model, however. Thus if all potential models provide a poor fit, there is still a winning model albeit it being of poor quality. To account for this drawback we developed a two-step algorithm to identify the main drivers of extreme wind speeds for the studied domain:

(a) By applying the AIC criterion we determined the best model for each grid cell checking combinations of all 9 computed leading EOF time series as potential predictors and $I_{k,y,9X}$ as the predictand. As $I_{k,y,9X}$ is sufficiently normally distributed a multi-linear regression model is a natural candidate. The simplest model would be a constant straight line; the most complex model would include all 9 predictors. The AIC step-wise algorithm in R (MASS package) examines every possible combination of the 9 predictors and estimates the model with the lowest AIC score, thereby providing the "best" combination of drivers for every grid cell.

(b) To determine the driver explaining the most variability, the predictors of the winning AIC model for each 5x5 grid cell are tested for significance using a t-test. Any predictor not significant at the 95% significance level is discarded from the AIC model. For each grid cell, the predictor associated with the largest regression coefficient is then considered to be the most important (winning) large scale driver for the respective grid cell. Due to the t-test constraint there can be grid cells where no significant driver can be determined.

This two-step algorithm is applied to both System 4 and ERA-Interim so that the internal drivers explaining the most variability of extreme winds can be compared between the seasonal forecast and reanalysis. This could help to understand why the potential predictability in seasonal forecasts is sometimes lower than in the real world as discussed in Eade et al. (2014), as we can examine what drives these extreme wind speeds (model internally). In case the extreme wind speeds are caused by different drivers in System 4 compared to ERA-Interim a potential lack of skill in System 4 for extreme wind speeds could be associated with the disparate generating drivers. Additional to the presentation of the winning driver for every grid cell we provide correlation maps of selected drivers to illustrate the different driving mechanisms between the seasonal forecast and the reanalysis.

4.3.5 A multi-linear regression model to statistically predict extreme wind speeds

The statistical model is trained to predict the $I_{k,y,9X}$ value for the upcoming DJF season based on the EOFs computed from MSLP ERA-Interim fields entailing data from October 15$^{\text{th}}$ to November 15$^{\text{th}}$ of every year. This period is centered on the initialization date of System 4 and thus imitates the lead time of 2-4 months of the

dynamical seasonal forecast model. As we include data until the 15th of November we do include more information in the statistical model than what would be available to System 4; however we balance that by including data all the way back to the 15th of October. Therefore we consider this as a fair base of comparison between the skill of the statistical and dynamical model.

The procedure of training the model is similar to the identification of the large scale drivers for the extreme wind speeds. The 9 leading EOF patterns for the October-November period are used as a pool of predictors to predict $I_{k,y,9X}$ for the coming DJF season. The model selection is equivalent to the algorithm described in the previous section: The selected predictors of the winning AIC model are tested for significance using a t-test. As $I_{k,y,9X}$ is sufficiently normally distributed the natural choice for the model is a multi-linear regression model again.

To account for a potential over-fitting of the model we performed a 2-fold cross validation of the winning model, thus each year is used as a training set and as a testing set once. In case no significant driver can be determined for a grid cell, there is of course no statistical model for that respective grid cell. The skill of the statistical model and the seasonal forecast model is examined by a correlation with the $I_{k,y,9X}$ value for all available grid cells of the ERA-Interim reanalysis data. Particularly the skill for Central and Northern Europe (as defined by the rectangular box in Figure 4.4.6) is analyzed.

4.4 Results

4.4.1 Differences in EOF patterns between System 4 and ERA-Interim

The 9 leading EOF modes of DJF MSLP data are presented in Figures 4.4.1 and 4.4.2 for ERA-Interim and ECMWF System 4, respectively. The comparison between the two EOF modes from EOF2 onwards generally has to be treated with care due to the degenerated subspace problem discussed earlier. Although the overall patterns look similar, the comparison with the patterns provided by NOAA reveals some differences for the EOFs calculated from ERA–Interim. One reason for that could be the slightly different definition of the MSLP anomalies as we are using 6–hourly data for the calculation of the EOFs whereas monthly values are used by NOAA. Using 6–hourly data naturally incorporates a higher variability into the EOF analysis.

Clearly the leading EOF for both the dynamical prediction model and the reanalysis represents the expected NAO pattern featuring the Icelandic Low and the Azores High. The location of this dipole however appears different for the two data sets. It seems that the centers of these two pressure patterns are more downstream in System 4 compared to ERA-Interim. As a result the pressure gradient over Europe appears smaller in System 4. The second EOF (PNA) also appears shifted downstream in System 4, reaching all the way into the European continent. Generally it seems that there appears to be a model bias that "smears" the variability patterns in a downstream direction.

Another example for that is the third EOF (Western Pacific (WP) pattern) which

also seems to explain a considerable amount of variability in the European region in System 4 although its center of action should lie in the Pacific. In System 4 it appears almost as a mirrored NAO with high pressure anomalies over Greenland and low pressure anomalies over the Central Atlantic, inconsistent with the reanalysis-based EOF patterns. The sixth (SCA) EOF in comparison exhibits a very similar pattern for both ERA-Interim and System 4 with a blocking situation over Central Europe. The eighth (EPNP) EOF with its dipole over the Pacific and a strong low pressure feature over the British Isles is also comparable between the two data sets. Other EOF patterns, however, look completely different as for example the fifth (EA) and the seventh (TNH) EOF.

Dawson et al. (2012) could show how the North Atlantic patterns of variability associated with the ECMWF model depend strongly on resolution. The dominant four patterns (NAO, EA, SCA, Atlantic ridge) only emerge in the natural variability of the model when higher than T511 resolution is used. This could be a partial explanation for why System 4 has such a different set of patterns of variability.

The explained variance by the different modes differs considerably between System 4 and ERA-Interim. Whereas the NAO accounts for around 20% of the inter-annual MSLP variability in ERA-Interim, it accounts for about 35% in System 4. More generally the explained variance (added up over all 9 modes) for the leading nine EOFs in System 4 ($\sim$87%) seems to be considerably larger than for ERA-Interim ($\sim$72%). There is a more abrupt change in explained variance in System 4 between the fourth (EAWR) and the fifth (EA) EOF, also. The decline of variance across the nine EOFs in ERA-Interim is smoother compared to that. As these

EOF patterns manifest so differently in the two data sets we decided only to use the EOF eigenvectors of ERA-Interim for calculating the time series and thus for identifying the drivers of extreme wind speeds as described in the Chapter 4.3.1.

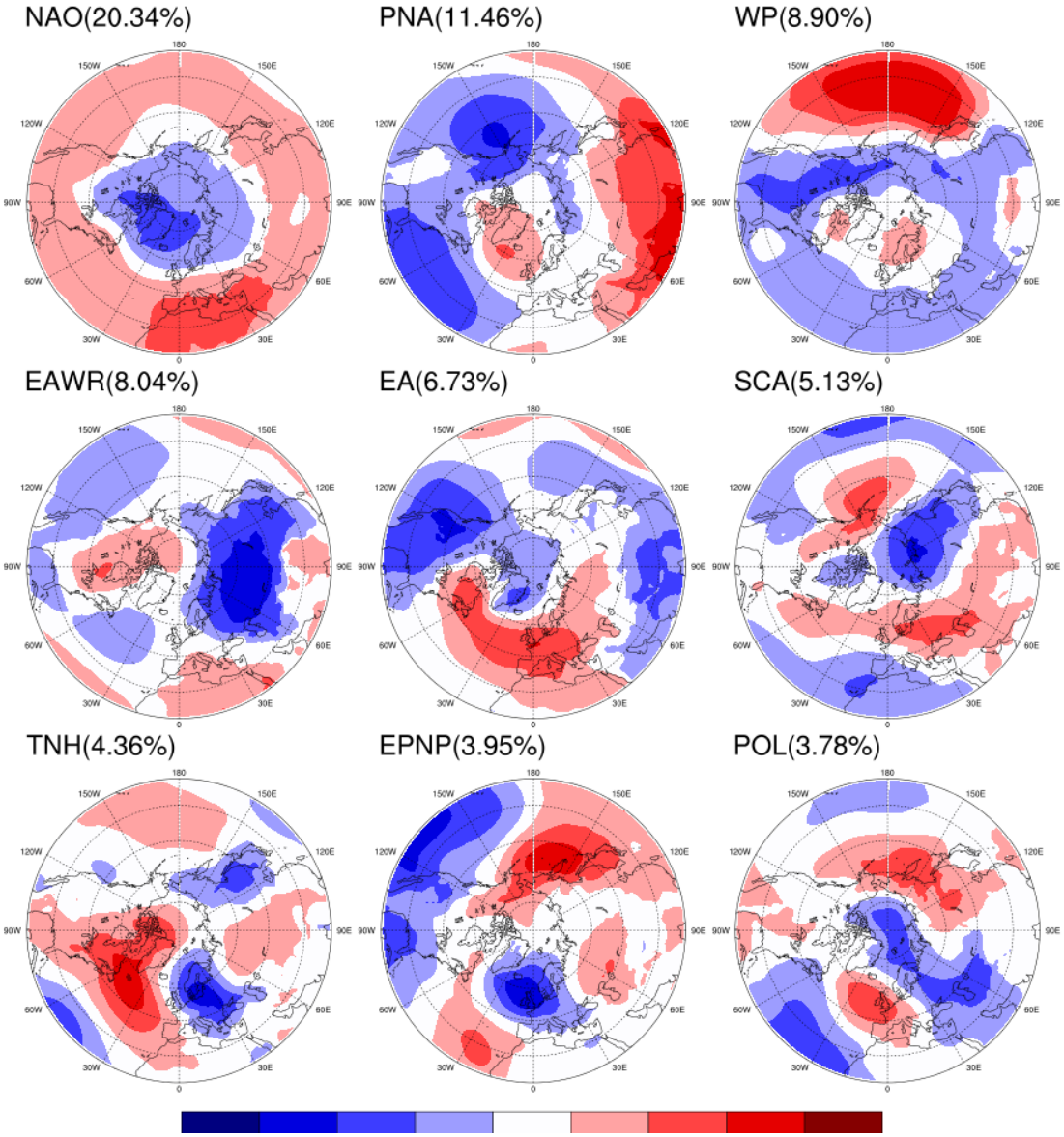


Figure 4.4.1: *The leading 9 EOF modes of MSLP during DJF in ERA-Interim. The explained variance of every mode is given after the conventional name of the EOF mode*

Figure 4.4.3 depicts the NAO time series of the EOF derived indices as well as the

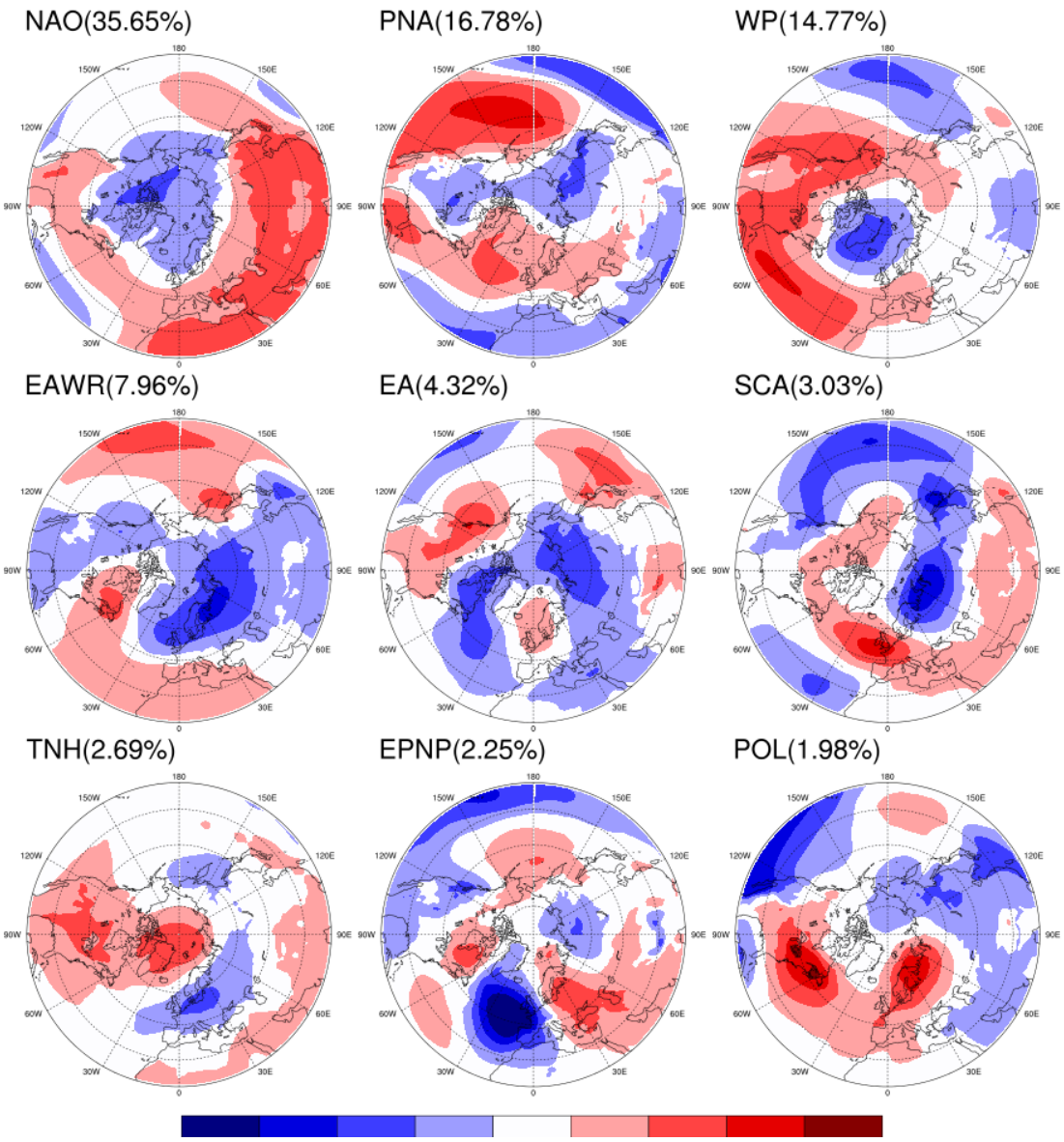


Figure 4.4.2: *The leading 9 EOF modes of MSLP during DJF in ECMWF System 4. The explained variance is given after the name each EOF mode would have if the NOAA nomenclature was used.*

grid-box-based indices. Note that the System 4 EOF based index uses ERA-Interim loading patterns and System 4 MSLP anomalies. The correlation between the two respective EOF-based NAO time series is 0.49 whereas the correlation between the grid-box-based time series is only 0.32. Evidently the EOF based and grid-box-based

NAO indices for System 4 appear to differ from each other considerably for certain years (e.g. 1984, 1999 or 2010) as well. The correlation between these two is 0.69 whereas the correlation between the two ERA-Interim indices is 0.80. The reason for the lower correlation of the System 4 indices could be due to the definition of the grid-box-based index using Lisbon and Reykjavik as reference locations is generally less suitable to assess the true MSLP pressure difference as the center of action of the NAO dipole is displaced in System 4 even when using the eigenvectors of ERA-Interim for its calculation. This could be similar to the findings of Ulbrich and Christoph (1999) who found that the center of actions of the NAO dipole shifts downstream for some GCM future climate projections.

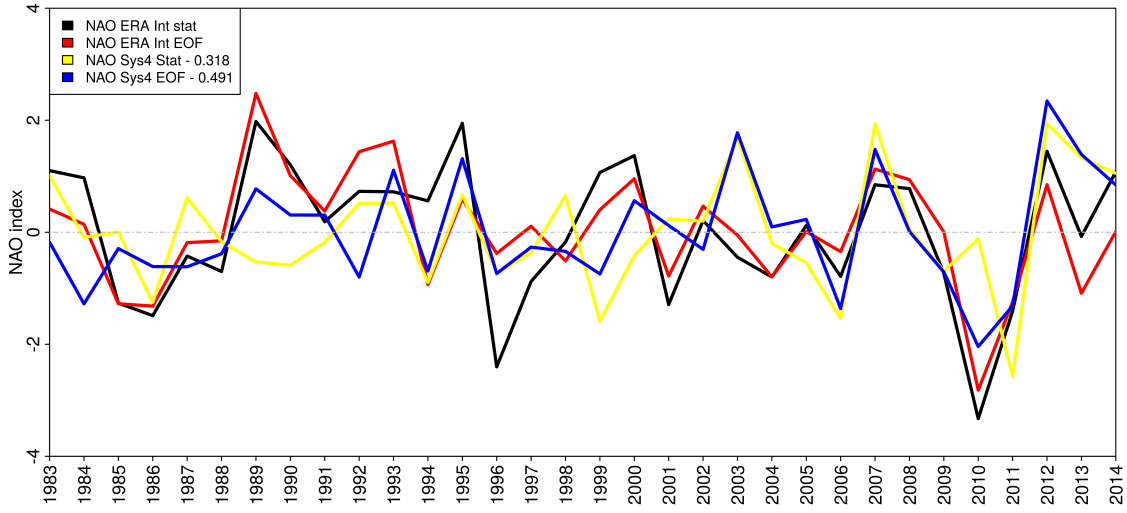


Figure 4.4.3: *Time series for the EOF based and the grid-box-based NAO index for System 4 and ERA-Interim. The correlation between the respective indices in ERA-Interim and System 4 is given in the legend. The years on the x-axis refer to the years of the January and February of a particular DJF season, thus 1990 refers to the winter 1989/90 etc.*

4.4.2 Drivers for extreme wind speeds in ERA-Interim and System 4

The MSLP EOF pattern featuring the largest (absolute) regression coefficient for every grid cell is presented in Figure 4.4.4. Each color plotted on the map represents one of the 9 leading EOF MSLP patterns using the nomenclature of the NOAA. As all indices bar the NAO index are not really comparable between reanalysis and seasonal forecast all indices except for the NAO are referred to by the number of the resulting EOF.

All predictors shown represent the "winning" coefficient for each grid cell, explaining most of the inter-annual variability of extreme wind speeds. Figure 4.4.4 depicts the drivers for $I_{k,y,95}$, however the maps for $I_{k,y,90}$ and $I_{k,y,98}$ look very similar. As expected from numerous studies (Pinto et al., 2009; Donat et al., 2010) the NAO explains the largest part of the variability of high wind speeds for the ERA-Interim reanalysis as it is known to be a key driver for North Atlantic storminess. The areas where the NAO has strongest effects is split by EOF5 which has also been identified as a variability pattern with a significant impact on European storminess in reanalysis products (Walz et al., 2018a or Mailier et al., 2006).

The predictors for the anomalies of System 4 projected on the ERA-Interim loading patterns, however, draw a somewhat different picture: While the northern part of the NAO dipole looks similar to the one observed for ERA-Interim, the southern part associated with the Azores High does not seem to be a driver of the inter-annual variability at all. Instead there is a mix of various drivers that seem to be responsible for extreme wind speeds for that area such as the EOF5 or the EOF7

pattern. The absence of the southern part of the NAO as a driver might be due to the downstream displacement of the high pressure associated with the Azores High in the leading EOF pattern (Figure 4.4.2).

The main driver for extreme wind speeds in Central Europe for System 4 appears to be the EOF6 pattern whereas it barely appears in ERA-Interim. The black crosses in Figure 4.4.4b denote grid cells in which the winning driver of ERA-Interim is at least amongst the top three drivers of System 4. This implies that the NAO might not be the most important driver for extreme wind speeds for the Central European region, however it still explains a considerable amount of variability in System 4 (see Figure 4.4.5). The differences in drivers between the two plots 4.4.4a and 4.4.4b can be explained by the different manifestations of the large scale MSLP patterns. When interpreting these results it has to be kept in mind however that the EOFs are constrained by their mutual orthogonality. This means that if the predictors for extreme winds are a linear superposition of two patterns that are not orthogonal, the EOF analysis will not yield these patterns. Furthermore the resulting EOF patterns can depend on the spatial domain (see Ambaum et al., 2001). One way to overcome this would be to use the effective regression weight, thus the linear combination of the EOFs weighted with the regression weights. The differences between the identified drivers for the reanalysis and the seasonal forecast become more obvious when comparing the correlation maps for the NAO and EOF6 variability patterns with extreme wind speeds individually. The correlation between $I_{k,y,95}$ and the NAO and the EOF6 for both ERA-Interim and System 4 are presented in Figure 4.4.5. Correlations significant at the 95% significance level are marked with a black cross. The correlation pattern of the NAO for ERA-Interim exhibits the prominent dipole

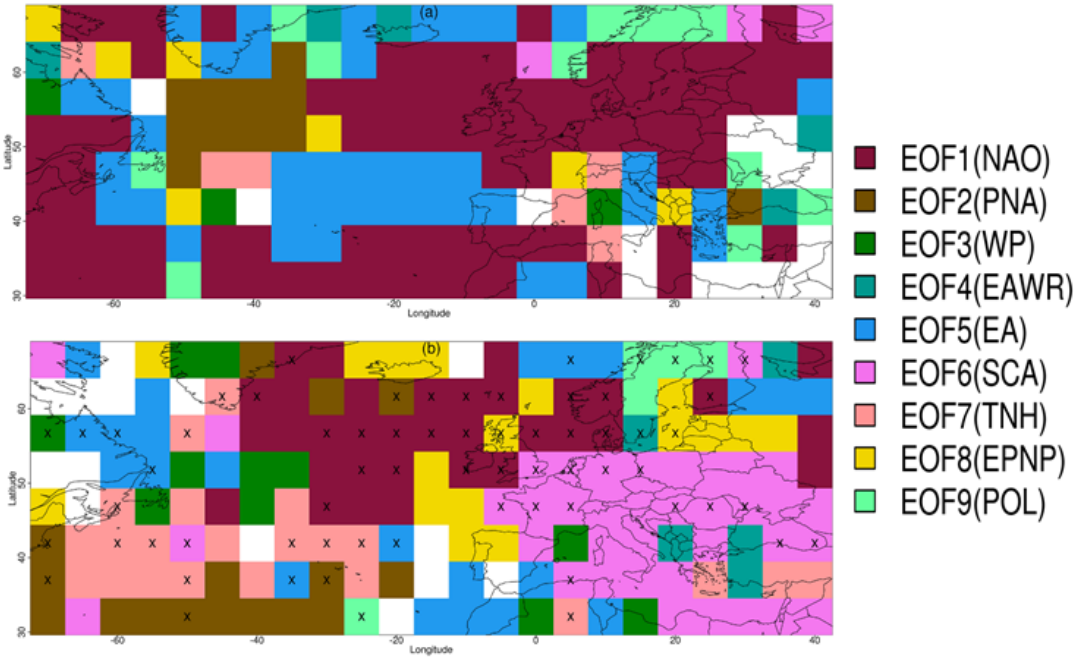


Figure 4.4.4: *MSLP variability drivers associated with the largest absolute regression coefficient for every grid cell for explaining the inter-annual variability of $I_{k,y,95}$. If no significant driver could be determined the grid cell remains white. Winning drivers for ERA-Interim are shown in the top figure (a), winning drivers for System 4 are shown in the bottom figure (b). The black crosses denote grid cells in which the winning driver of ERA-Interim (a) was at least amongst the top 3 for System 4. The nomenclature for the EOFs is according to the NOAA standards for the NCEP reanalysis.*

pattern over the North Atlantic featuring positive correlations between the NAO and extreme wind speeds over Northern Europe. Negative correlations are present over Southern Europe and over the Atlantic between 30°-40°N.

The correlation patterns look considerably different for ERA-Interim and System 4. Whereas the northern part of the correlation for the NAO looks similar in both data sets, the southern part barely shows any significant negative correlation in System 4. The difference in correlations for the EOF6 pattern is even more striking. Basically the entire European continent features a negative correlation between the EOF6

(SCA) time series and the inter-annual variability of extreme wind speeds in winter in System 4, whereas the main region of correlations in ERA-Interim appears to be (compliant with its name) the Scandinavian region. Also the correlation pattern features different arithmetic signs. As the correlation between the EOF6 and extreme wind speeds in System 4 over Central Europe is fairly strong (< -0.5) it explains why the EOF6 appears as the most important driver of the variability as seen in Figure 4.4.4. For the sake of completeness we present the results for model specific results (utilizing System 4 eigenvectors in the Supplementary Material). There the first EOFappears only occasionally as the most important MSLP variability pattern, whereas the PNA (second EOF), the WP (third EOF) and the EPNP (eighth EOF) appear as the patterns explaining the most variability of extreme wind speeds in the North Atlantic area. The shifted importance of the different patterns appears to be related to the observed downstream displacement of the center of actions of the aforementioned MSLP variability patterns (Figure 4.4.2) as for example the second EOF (PNA) which explains variability in the European domain as well.

4.4.3 Predictability of high wind speeds in the ECMWF System 4

The estimated predictability (using mean PP) of $I_{k,y,9X}$ is presented in Figure 4.4.6 for all three evaluated percentiles. The figures show the average PP for every 5x5 grid cell over the period from DJF 1982/83 until DJF 2013/14. For exceedances of all three percentile thresholds there seem to be three main pockets of large PP values: Over the Scandinavian region, over the Central-Eastern Atlantic and over Eastern Canada/Newfoundland. Interestingly two of these areas are towards the edge of the analyzed domain.

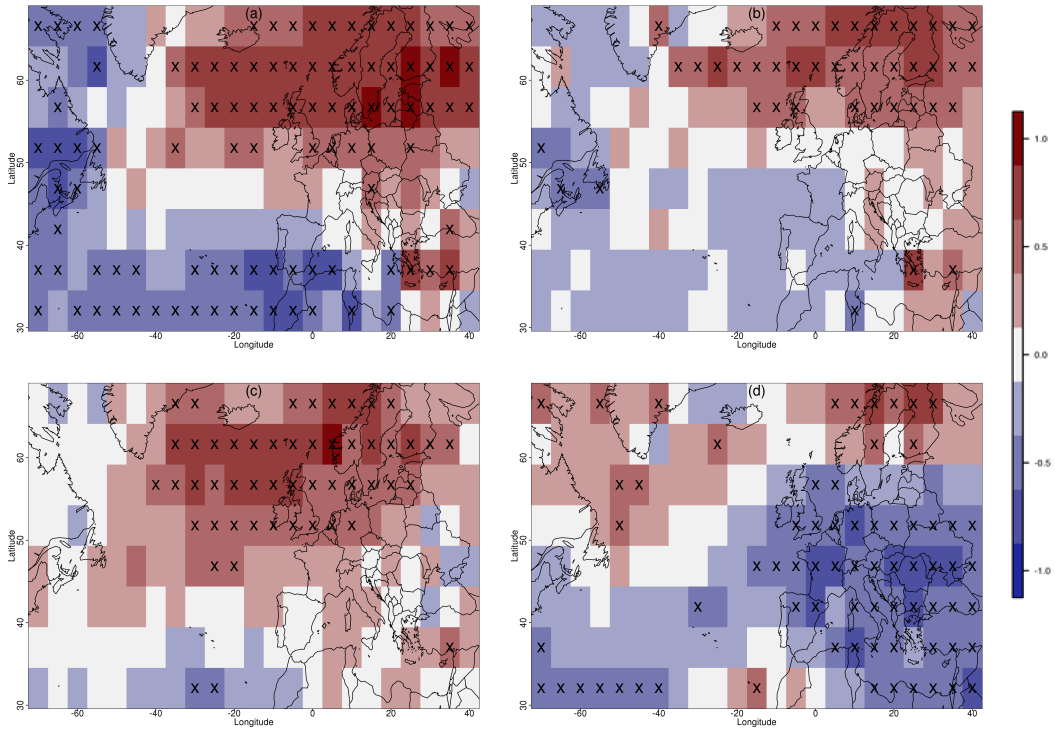


Figure 4.4.5: *Correlation maps between EOF1 (NAO) and $I_{k,y,95}$ time series for ERA- Interim (a) and System 4 (c) and between EOF6 (SCA) and $I_{k,y,95}$ time series for ERA-Interim (b) and System 4 (d). All significant correlations at the 5% confidence level are marked with a black cross*

The maximum predictive power across all three percentile exceedances is found in the Scandinavian region. This could be due to the fact that the prior uncertainty in these regions is relatively large as the climatology here exhibits a strong year-to-year variability. This indicates that the initialized runs provide useful information to narrow this spread for the upcoming winter season. The small climatological spread is possibly also the reason for the fact that there is no PP over the Atlantic within the classic storm track. The prior uncertainty in this region is relatively small in this region and so the forecast does not provide any more useful information or is able to reduce the uncertainty respectively.

The PP does generally not decrease with increasing percentile threshold. The overall pattern looks very similar in all three plots. Interestingly certain areas feature even higher PP values for the higher percentiles (e.g. Northern Atlantic west of the British Isles or across Central Europe). Generally System 4 appears to provide additional information to the climatology where the amount of windstorms associated with extreme winds has a larger inter-annual variability. This makes sense as the climatological distributions in these areas will be reasonably broad, thus the prior uncertainty is large. From an end-users' point of view this is very useful information as it seems that the forecast of high wind speeds exhibits a good degree of sharpness in areas where potential damage could occur from those winds.

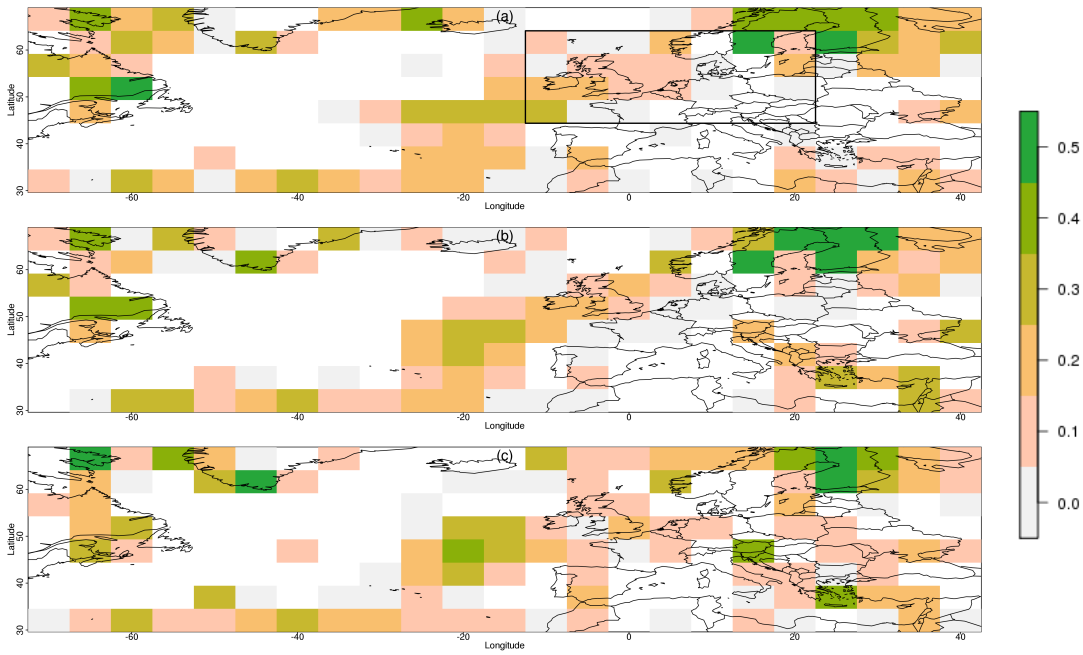


Figure 4.4.6: *Mean Predictive Power (PP) of ECMWF System 4 for the DJF seasons from 1983/84 to 2013/14 for $I_{k,y,90}$ (a), $I_{k,y,95}$ (b) and $I_{k,y,98}$ (c) The rectangular box marks the regions for which regional average time series are analyzed (c.f. Figure 4.4.8).*

4.4.4 Evaluation of the statistical model in comparison to System 4

The statistical model developed in Chapter 4.3.5 is based on the 9 leading EOFs of ERA-Interim mid-October to mid-November MSLP data. That way it also attempts to predict winter storminess based on October to November initial conditions and ensures a fair level of comparison between the statistical model and the dynamical seasonal forecast product System 4. Figure 4.4.7 shows maps of the correlation between the respective model and ERA-Interim for the normalized sum of extreme winds $I_{k,y,95}$ during DJF.

Compared to System 4 the statistical model undoubtedly shows more grid cells with a significant correlation (skill) for predicting the extreme wind speeds during the DJF season. The significant grid cells for System 4 tend to be at the edge of the studied domain, whereas the major part of grid cells with significant correlations for the statistical model appears over Central and Northern Europe featuring correlations up to 0.5. Interestingly the correlation for System 4 is highest where there was also the highest magnitude of PP (e.g. Scandinavian region in Figure 4.4.6) which confirms that the PP can be used as an alternative tool to examine the predictive skill of an ensemble forecast. A definite drawback to the statistical model is the fact that there is not necessarily a statistical model for every grid cell. In case none of the 9 EOFs of the November MSLP turns out to be significantly related to local extreme wind speeds, no model can be determined. The maps look very similar for $I_{k,y,90}$ and $I_{k,y,98}$ so the result is not sensitive to the wind speed threshold. Therefore we choose not to show them at this point.

Interestingly many grid cells for which there is no statistical model exhibit some skill in the dynamical forecast. An interesting approach could be to combine the two forecast models. One possibility would be to check for skill in System 4. If none can be found the statistical model could be checked for a significant model in that particular grid cell. If skill can be found that grid cell could be used from the statistical model. Judging from Figure 4.4.7 this "hybrid" model would exhibit skill in larger areas than any individual model.

To include all of the previously discussed predictability scores and correlations into one plot all of these items were averaged over the defined rectangular box (see Figure 4.4.6) including large parts of Central Europe and the British Isles, thus roughly the area which is most affected by European windstorms (e.g. Leckebusch et al., 2008a). The time series (Figure 4.4.8) of $I_{k,y,95}$ in ERA-Interim nicely captures the inter-annual variability of extreme winds in Central Europe with peaks for the stormy seasons 1989/90, 1992/93, 1999/2000, 2006/07 and 2013/14 as these featured some of the most prominent European windstorms in the last 30 years. The minima are found for the winters 1984/85 and 2009/10. Both of these seasons also exhibited strong negative values of the NAO (see Figure 4.4.3). It seems that the ensemble spread of System 4 is approximately as big as the yearly variations of ERA-Interim; however the mean of the ensemble shows very little variability over the entire time period. This small signal variability is a common feature when taking the mean of a large ensemble size as it is provided with the 51 members by System 4.

The correlation between the ensemble mean and ERA-Interim of 0.19 is not significant at the 95% significance level. Every winter season in the System 4 ensemble

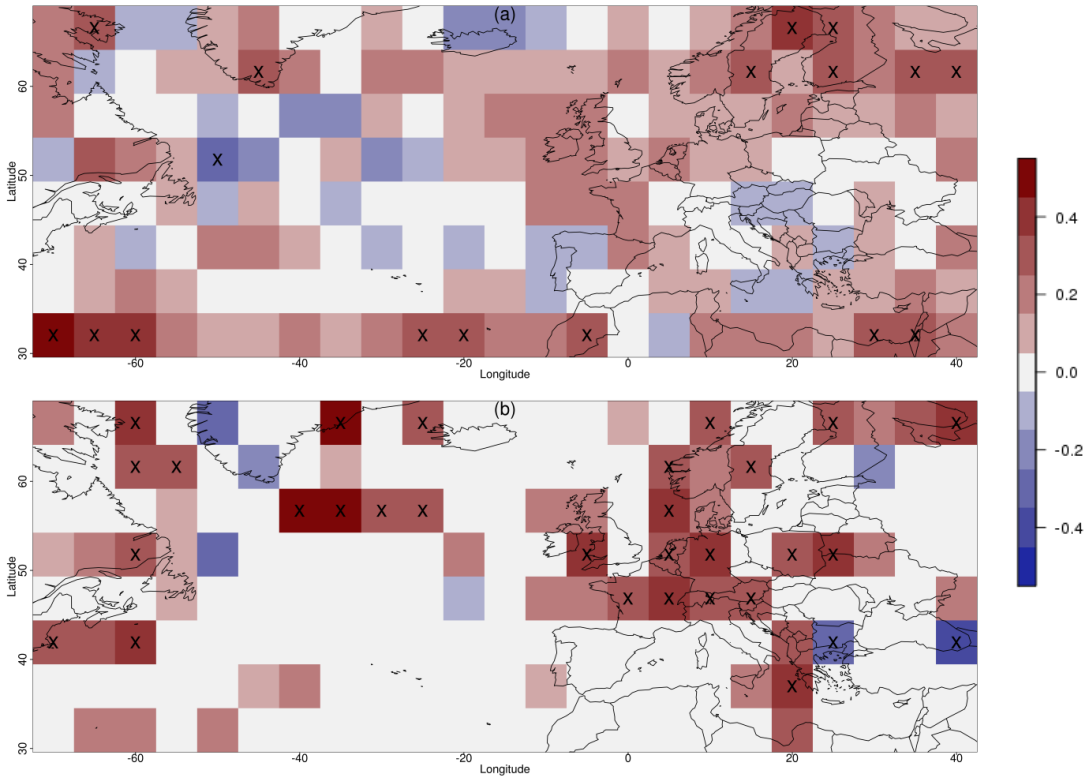


Figure 4.4.7: *Correlation maps for $I_{k,y,95}$ time series between ERA-Interim and the ensemble mean of System 4 (a) and ERA-Interim and the statistical model based on November MSLP data (b). All significant grid cells at the 5% confidence level are marked with a black cross.*

mean appears to be of a similar intensity whereas the observations exhibit a strong year-to-year variability. Compared to the range of the entire ensemble, the inner 50th percentile appears sharper. This leads to the conclusion that many ensemble members end up in the same state, thus agree well to one another. The time mean of $I_{k,y,95}$ over the entire time is very similar for System 4 and ERA-Interim. This implies that the mean of the ensemble forecast manages to get the mean extremeness of wind speeds correct over the 32 years, however fails to predict the inter-annual variability. The inter-annual variability in the statistical model is considerably larger. In addition it correctly predicts some of the peaks (e.g. 1999/2000 or 2011/12) al-

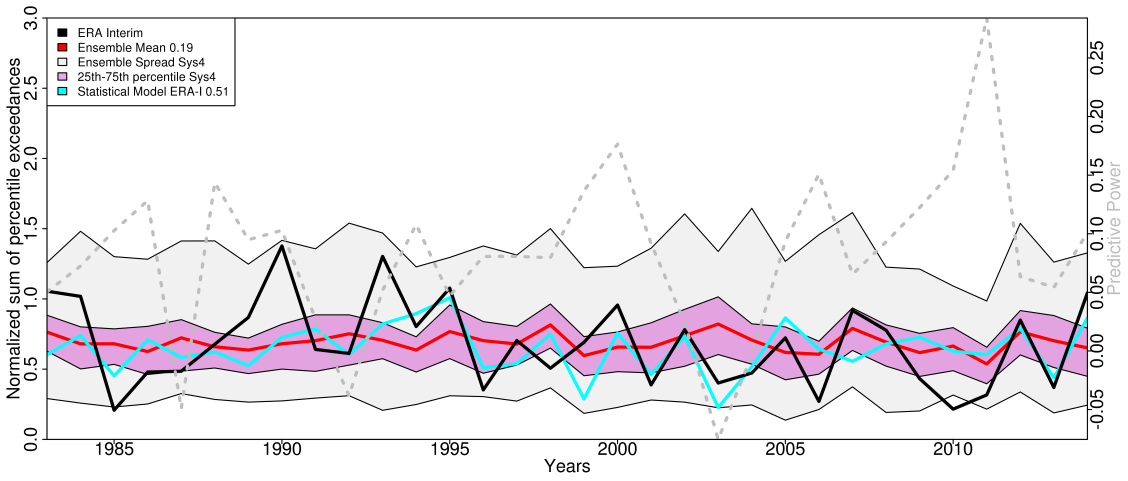


Figure 4.4.8: *Time series of $I_{k,y,95}$ spatially averaged for the European region as defined in Figure 4.4.6. The black line corresponds to ERA-Interim, the red line corresponds to the mean System 4, the purple band depicts the inner 50^{th} percentile of System 4 and the grey band depicts the range of the entire ensemble of System 4. The light blue line represents $I_{k,y,95}$ as predicted by the statistical model. The dashed light grey line depicts the PP whose y-axis is plotted on the right hand side of the plot. The correlations between the ensemble mean of System 4, the ensemble median of System 4 and the statistical model with ERA-Interim respectively are given in the legend after the respective label. The years on the x-axis refer to the years of the January and February of a particular DJF season, thus 1990 refers to the winter 1989/90 etc.*

though it fails to predict the prominent winter of 1989/90. The statistical model occasionally predicts more extreme seasons than ERA-Interim (2001/02 or 2009/10). The correlation of 0.51 is significant at the 95% significance level.

Generally it seems that unusually low extreme wind seasons like 1984/85 and 2009/10 are harder to predict for both the dynamical and the statistical model. The predictive power is negatively correlated (-0.5) with the spread of the ensembles which is in line with the assumption that less spread/uncertainty in the forecast is related to a higher predictability (information gain) of the extreme wind speeds. Despite

being less uncertain, the ensemble members appear to agree more with each other than with the actual observation. This compares to an underconfident forecast as discussed in Eade et al. (2014). In other words there are barely any seasons in which the ensemble mean forecast can make a prediction of the actual intensity (peak intensity) of the season. In terms of information gain, however, the seasonal forecast still provides information additional to the climatology for some years; in particular for the winters 1999/2000, 2006/07 and 2010/11 (see the peaks of the PP curve in Figure 4.4.8).

4.5 Summary and Discussion

This study utilized the ECMWF System 4 and ERA-Interim to advance the understanding of the large-scale drivers associated with extreme winds (i.e. associated with European winter windstorms) and their predictability. The NAO (first EOF) pattern in the seasonal forecast seems to be shifted downstream so that its center of action is relocated towards the East. This downstream displacement of variability modes is also observed for the second EOF (corresponding to PNA in the reanalysis) and the third EOF (WP), which are extended into the Atlantic.

The correlation of the two EOF-based NAO time series between the reanalysis and the projected System 4 index is around 0.49 whereas it is only 0.32 for the grid-cell-based index. The lower correlation for the latter could potentially be explained by the shifted location of the EOFs in System 4, similar to Ulbrich and Christoph (1999) who argued that the rigid location for the grid-cell-based NAO index could be inaccurate for some future GCM predictions (ECHAM4+OPYC3 in their case) as the center of action of the NAO is displaced downstream. A similar displacement

is observed for the seasonal forecast. The correlation skill of the NAO in System 4 is in good agreement with the findings of O'Reilly et al. (2017). They computed a correlation of 0.31 for the NAO index derived from 500 hPa geopotential height data using the (atmosphere only) seasonal experiments for the entire 20th century Weisheimer et al., 2017) and of around 0.40 for the 31 years between 1980 and 2010, thus close to our estimated correlation of 0.49. The difference might be partly due to the slightly different time period used in their study.

The question about the predictability or uncertainty of extreme wind speeds in System 4 was approached with the statistical entropy and the Predictive Power (Schneider and Griffies, 1999; Tang et al., 2007) in particular. Evidently there is more information in the seasonal forecast (larger PP value) in areas where the variability of the climatology is fairly large. Thus the seasonal forecast is able to reduce the uncertainty of the forecast distribution compared to the climatological forecast. This applies especially for some regions in the Central East Atlantic, Eastern Canada and northern Scandinavia.

The variability in the climatology of extreme winds within the prominent storm track is comparatively low so that the seasonal forecast cannot provide much additional information for these areas. Overall, the PP value proves to be a valid measure of uncertainty quantification. Fernández-González et al., 2017 investigated the uncertainty of wind speeds for the ECMWF EPS operational weather forecasting ensemble with a predictability index that evaluates the average inter-quartile range for a 30 day period against the current forecast. In a way this is similar to the PP as it estimates in how far the uncertainty of the climatology can be reduced by

a forecast, thus whether or not the predictability is better than "usually". Overall the conclusions they draw from their analysis of the ECMWF EPS are in line (see below) with System 4 although the forecasts are on different time scales and in their case on a much smaller spatial scale.

The map of the physical drivers of extreme wind speeds for System 4 looks fairly different to the reanalysis. The NAO acts as the dominant driver in the North Atlantic, however it does not seem to explain much variability in the Central Atlantic. The sixth EOF (SCA) is identified as the main driver for extreme winds over Central Europe. When the model specific eigenvectors are used to produce the time series, the differences between ERA-Interim and System 4 are even more striking. It has to be kept in mind here that we used the EOF patterns of ERA-Interim to create the EOF time series for System 4. By doing that we could ensure a fairer comparison with regards to the selection of drivers, however the question remains why the EOF patterns look fairly different between System 4 and ERA-Interim. Physically a system generating extreme wind speeds (i.e. an extreme cyclone in the North Atlantic) should be the same within both System 4 and ERA-Interim.

The actual EOF patterns shown in Figure 4.4.1 and 4.4.2 should in fact affect the probability of occurrence of these large scale systems. We therefore cannot make any assumptions on how often the large-scale conditions generating windstorms in ERA-Interim actually occur in System 4 and how much variability of extreme wind speeds they can explain.

Even though the NAO in ERA-Interim and System 4 are sufficiently well correlated

as shown in this study and also by Kim et al. (2012) or Weisheimer et al. (2017), the System 4 internal NAO fails to act as the most dominant driver associated with the inter-annual wind speed variability in Central Europe. This would suggest that the skillful prediction of the NAO, which is certainly possible (as shown e.g. with GloSea5 by Scaife et al., 2014), may be used to deduce the winter storminess in the real world building on empirical relationships between NAO and storminess. However, it seems that the NAO and the winter storminess in the model itself are not as related as in the real world.

On the one hand this would mean that extreme wind speeds in the European region are physically associated with different variability patterns even if the large scale wind speed variability pattern looks correct "being right for the wrong reasons". On the other hand this also means that, by improving the model-internal MSLP variability, the skill of the seasonal forecast could potentially be substantially increased. Of course this would not be constrained to the skill of extreme winds but may also affect the skill of various other atmospheric variables.

The spatially averaged time series revealed a major inconsistency of System 4 in comparison to the ERA-Interim reanalysis: Whereas the time mean over the 32 year period of System 4 and ERA-Interim agree well, the year-to-year variability of extreme winds is clearly underestimated by System 4. This lack of variability is a common feature of an ensemble mean which is to be expected however: We would not expect the ensemble mean to be a good forecast (in a deterministic sense) of the atmospheric state more than a month later. The expectation is that the ensemble mean is closer to the climatology of the model than most of the individual members.

This is seen clearly in the time series of the ensemble mean. Therefore, the fact that the System 4 ensemble mean does not predict anomalous seasonal averages in ERA-Interim is to be expected and does not imply failure to predict. The primary purpose of the ensemble is to make a probabilistic prediction of the state for the next season.

The spread of the inter-quartile range of System 4 appears fairly sharp, implying that at least half of the ensemble members predict wind speeds relatively close to each other. This is in accordance with Fernández-González et al., 2017 who determined the inter-quartile range as the most balanced uncertainty quantification for wind speeds over Spain in the ECMWF EPS forecasting system. They also note that the ensemble mean should be accompanied by some measure of uncertainty. Based on Figure 4.4.8 we can confirm this claim.

While the System 4 prediction skill for extreme winds is relatively low, possible avenues for improvements may be to use multi-system approaches. For example, Athanasiadis et al. (2017) could show that a multisystem of seasonal forecasts consisting of GloSea5, the National Center for Environmental Prediction (NCEP) CFS version 2 CFSv2 (Saha et al., 2014) and the CMCC Seasonal Prediction System 1.5 (Materia et al., 2014) has unprecedented high predictive (correlation) skill for the NAO and AO of 0.85 for the short period of 1997-2014. They follow the arguments by DelSole et al. (2014) who found that the enhanced predictive skill of a multisystem of forecasts exceeds the expected increase due to larger ensemble sizes alone.

DelSole et al. (2014) argue that the increase in predictive skill is more consistent

with the addition of new signals from the different forecast systems. Thus one way to possibly improve the forecasting skill would be to incorporate more than one seasonal forecast product for future research. Additionally to assess predictability in a probabilistic sense one would need to look at for example a tercile forecast of the NAO rather than just the ensemble mean. That way the deviation from a neutral NAO index for example could be investigated.

Befort et al. (2018) showed significant skill in their event-based study analyzing the predictability of European windstorm occurrence in seasonal forecasts. An explanation for the higher skill in their study could be that the inter-annual variability in the frequency of windstorms in the North Atlantic is less affected by the variability of extreme wind speeds at a particular grid cell. The tracking algorithm used to identify windstorms is based on an exceedance of a local percentile. The magnitude of the exceedance of the percentile however is irrelevant for the identification of an event, whereas it is relevant for our $I_{k,y,9X}$ value. Thus, the degree of exceedance adds an extra dimension of uncertainty to the skill assessment.

It has to be noted, also, that their analysis only comprises the years 1992-2011 and utilizes wind speeds in 925 hPa whereas we use 10-meter wind speeds over the period 1982-2013. We did some brief analysis only taking into account this shorter time span and found correlations of similar magnitude and pattern to what they had found in their paper, e.g. off the coast of France and the UK in particular (not shown). Whereas their results appear significant, our map of correlations for the period from 1992-2011 still shows very little statistical significance. The finding of lower skill for the 1982-2013 period is also in line with their findings.

The inter-annual variability is captured more accurately by the statistical model, however the major storm seasons (e.g. 1989/90) are not well predicted either. There is a negative correlation between the PP and the spread of the ensemble which is in line with the expectation that the PP is particularly large when the uncertainty in the forecast is low, implying that there is additional information in the forecast compared to the climatology.

The two main conclusions of out presented study here are:

- The EOF patterns of System 4 look considerably different compared to the reanalysis. Even when projecting the MSLP anomalies of System 4 onto ERA-Interim loading patterns, the large-scale drivers correlated with extreme winds in System 4 are fairly different to what is observed in the reanalysis. This could be one of the reasons for the low skill of the ensemble mean with regard to the observed value of $I_{k,y,9X}$. The low skill is in accordance with Della-Marta et al. (2010) who found no skill for the 95^{th} percentile of wind speeds in the predecessor of System 4 for months past the first lead month (lead months 2-4). The open question here remains why the coupled model that runs freely after initialization produces such different patterns of atmospheric variability compared to the reanalysis that is constraint by observations. Especially as the generating systems of extreme winds in DJF (extreme windstorms) should be the same in both System 4 and ERA-Interim.

- A cross-validated multi-linear regression model using large scale MSLP patterns from mid-October to mid-November can provide significant skill for the upcoming DJF season with regard to extreme wind speeds. Thus, October

and November MSLP data contain valuable information regarding the upcoming DJF storminess for vulnerable regions, pointing to potential predictability in the climate system that may not be captured by the numerical model in System 4. The statistical model could potentially be further improved by incorporating longer time series, e.g. by using ECMWF ERA-20C (Poli et al., 2016).

For future research it would be interesting to investigate whether there is a temporal progression in the spatial shift in System 4 compared to ERA-Interim from initialization to increasing lead times, e.g. carry out an EOF analysis for different lead times. Additionally the identification of drivers could be applied to other variables, especially to those which have already been validated to feature higher skill (e.g. winter temperature, Kim et al., 2012 or Ogutu et al., 2017) than the erratic wind speeds. The statistical model developed in this study provides a good indication about the potentially predictable storminess of the upcoming winter season and could provide a useful tool for the impact/actuarial community.

Chapter 5

Spatial Variability and potential maximum Intensity of storms over Europe

"The Westerly Wind asserting his sway from the south–west quarter is often like a monarch gone mad, driving forth with wild imprecations the most faithful of his courtiers to shipwreck, disaster, and death."

– Joseph Conrad, British–Polish writer

Abstract

Extra–tropical wind storms pose one of the most dangerous and loss intensive natural hazards for Europe. However, due to only 50 years of high quality observational data, it is difficult to assess the statistical uncertainty of these sparse events just based on observations. Over the last decade seasonal ensemble forecasts have become indispensable in quantifying the uncertainty of weather prediction on seasonal time scales. In this study seasonal forecasts are used in a climatological context: By making use of the up to 51 ensemble members a broad, physically consistent statistical base can be created. This large sample can thus be used to assess the uncertainty of extreme wind storm features such as intensity or severity more accurately. In particular return periods and even a potential maximum intensity of windstorms and extra–tropical cyclones (ETCs) can be calculated depending on a specific cluster or region in Europe. A 100–year event minimum core pressure in Central Europe, for example, is estimated to be around 940 hPa, whereas it would be around 928 hPa for the British Isles. By employing extreme value statistics a potential minimum core pressure (maximum curvature) can be estimated as well. This is way below (above) a 1000–year event however, it can therefore be seen more as a physical barrier than a realistic scenario.

The following chapter is an edited version of the submitted NHESS manuscript: *Walz, M. A., & Leckebusch, G. C. (2018). Spatial variability and potential maximum intensity of winter storms over Europe, NHESS-2018-309, submitted on 23 October 2018*

This study was designed by Michael Walz. All data processing and analysis was carried out by Michael Walz. The clustering algorithm based on Gaffney et al. (2007) was implemented in FORTRAN by Michael Walz. The manuscript was written by Michael Walz. Gregor Leckebusch gave valuable comments which helped to bring out the added value of this study more clearly. This study is currently (March 9, 2019) under review.

5.1 Introduction and Motivation

European winter windstorms are responsible for extreme surface winds and heavy precipitation events that result in major flooding in many parts of the European domain. As a result windstorms have a vast impact on socio-economic structures of the resident societies. The storm Xaver that hit Europe in December 2013 was responsible for economic losses somewhere between 700 Million and 1.4 Billion Euros (AIR Worldwide, 2013) and around 10 casualties. The extra-tropical cyclones (ETCs) instigating these windstorms have their origin over the Northwest Atlantic. They usually follow an eastward trajectory until they eventually affect the European continent. ETCs play a crucial role in the reduction of the meridional temperature gradient as they convert potential energy into turbulent kinetic energy (Leckebusch et al., 2008b). This implies that ETCs act as a central nexus between the large scale dynamics of the atmosphere and direct local impacts, manifested in economic losses caused by associated extreme surface winds (Gaffney et al., 2007). Recent studies have shown that ETCs are steered by a variety of large scale drivers which to some degree can be used as a prognostic tool to estimate the amount of windstorms per winter season (Walz et al., 2018a, Vitolo et al., 2009, Mailier et al., 2006). This in turn means that a better understanding of the large scale variability and the properties of ETCs in general can have an enormous societal impact.

A general technique to categorise meteorological data and ultimately link it to large scale dynamics is represented by a clustering approach which has frequently been used in the literature (e.g. Philipp et al., 2007 or Leckebusch et al., 2008b). Thus, data is assigned to different clusters so that each respective cluster contains alike data. These clusters can then be investigated individually, thus individual prop-

erties can be calculated. If, for example, one cluster features systematically more intense storms than another cluster, this information could be used in a forecast mode. Hence, if a windstorm in a forecast exhibits properties of the intense cluster, assumptions could be made about the potential impact of the storm. In addition to the forecast mode, the properties of individual clusters can help to understand dynamics behind ETCs. One cluster for example could entail extreme secondary cyclones like windstorm Lothar in 1999 (Ulbrich et al., 2001)

In terms of ETCs there have been a handful of studies that have applied a clustering approach.Leckebusch et al. (2008b) used a k-means approach to classify meteorological circulation regimes which are accountable for windstorms in Europe. They could identify 4 principle circulation patterns (Primary Storm Clusters; PSC) that are responsible for the occurrence of harmful ETCs during the extended winter season (Oct-Mar). Jointly these 4 identified clusters contain more than 70% of the historic storms between the years 1958 until 1998. Blender et al. (1997) also used a k-means clustering approach; however instead of circulation patterns they classified the trajectories of identified ETCs. The major drawback of the k-means approach in the context of clustering is its requirement for the data to be of the same length. As windstorms/ETCs occur in different spells, this approach is only partly useful for the sake of the problem. Gaffney et al. (2007) proposed a probabilistic clustering approach in which ETCs are considered in a Lagrangian view as each of them features a unique track and life cycle. Their approach allows for trajectories to be assigned to clusters regardless of their duration. As this approach will be implemented for this study, a more detailed explanation follows in the Methods section of this paper.

Owing to new Reanalysis products like the European Centre of Medium-range Weather Forecasts (ECMWF) ERA-20C (Poli et al., 2016) or the NOAA 20CR (Compo et al., 2011) there is little more than 100 years of high quality windstorm "observations" on grid cell level. In order to estimate the uncertainty of high impact windstorms in terms of frequency and severity however, the amount of data is still too sparse to produce reasonable confidence intervals. Similar to Osinski et al. (2016) in which the ECMWF Ensemble Prediction System (EPS) is used as a data archive for creating a windstorm catalogue, this study approaches the retrospective predictions of ECMWF Seasonal Forecast System 4 (Molteni et al., 2011) as an archive of potential windstorms. Clearly none of the windstorms found in these forecasts ever happened, however each of them represents one possible physical consistent realisation of a potential reality. Due to the 51 ensemble members of System 4 this leads to a substantial increase (around 1500 years) in the available sample of potential extreme events. This will allow for a more accurate estimation of uncertainties regarding features of extreme windstorms, e.g. intensity or duration. The ensemble members are treated as statistically independent since studies have shown that there is little forecast skill for very high local wind speeds ($>98^{\text{th}}$ percentile; Walz et al., 2018b). As predictability indicates statistical dependence of the ensemble members, the inverse is also true; hence no/little predictability implies statistical independence of the original ensemble (DelSole and Tippett, 2007, DelSole, 2004).

The novel approach of this study is to gain a better understanding of the hazard uncertainty of windstorms and ETCs by utilizing the ensembles of seasonal retrospective forecast data in a climate archive approach. Events will be identified based on two different objective tracking algorithms, one of which is based on wind

speed whereas (Leckebusch et al., 2008a or Kruschke, 2015) the other one is based on maxima in curvature of the mean sea level pressure (MSLP) field (Murray and Simmonds, 1991). The events will be classified based solely on the shape of their respective trajectory. The different clusters will be analysed with regard to storm features such as maximum intensity, duration and celerity. Eventually a most and least intense storm cluster will be identified. Chapter 5.2 will describe the ECMWF System 4 data which is used for this study. Chapter 5.3 summarizes the method of the probabilistic clustering approach proposed by Gaffney et al. (2007), before Chapter 5.4 presents the results of this approach. The paper will close with a Summary and Discussion in Chapter 5.5.

5.2 Data

The idea of this study is to utilize the ECMWF System 4 (Molteni et al., 2011) as a climate archive in a way of assuming the 51 members of retrospective forecasts each resemble an artificial reality. System 4 was the operational seasonal forecast system until November 2017. This study uses retrospective forecasts which are initialised at November 1st each year from 1983 until 2013. Every run lasts for 7 months so that data from November 1st until May 31st are available. Due to the spin-up (avoiding potential "real storms" at the beginning to guarantee statistical independence) of the model and the focus on European winter windstorms only, the months December until March are used for this study. The ensemble entails 51 members which, combined with 31 years of data, is equivalent to 1581 "virtual" years or winter seasons respectively. That way the ensemble serves as a unique data archive which can be used to assess the statistical uncertainty more precisely compared to exploiting observational reanalysis data for this kind of estimation. System 4 is provided on

a spectral resolution of T255 which is the same resolution used for the Reanalysis ERA Interim. The perturbed initial conditions are produced using singular vectors and an ensemble of ocean conditions of the ocean model NEMO (Madec, 2008). The forecast system is based on the IFS cycle 36r4 of the ECMWF.

Two different types of identification methods are used for this study: Windstorms are identified using an objective windstorm tracking algorithm that is based on the exceedance of the local 98th percentile of wind speeds (Leckebusch et al., 2008a; Kruschke, 2015). Cyclones are identified using the cyclone identification method developed by Murray and Simmonds (1991) which is based on finding maxima in the curvature (c) of the MSLP field. Both times the tracking is carried out for the entire Northern Hemisphere, however only events affecting specific countries/areas of Europe are analysed within this study.

Due to the abundance of windstorms and cyclones it is possible to determine the hazard uncertainty of windstorms on a country/region level. This is implemented by only taking into account windstorms that affect a country at least once in their lifetime, i.e. by defining a radius around a country/an area through which a windstorm or a cyclone has to pass. Thus tracks, that never crossed a respective area (e.g. British Isles) are discarded from the analysis The area of the maximum wind field of a cyclone is usually found southeast of the core of the cyclone (Fig.1; Leckebusch et al., 2008a). That is why the selection radii for the tracked cyclones are slightly different to the windstorm ones, i.e. shifted a bit towards the northwest. As the cyclones identified by the Murray and Simmonds algorithm are not necessarily extreme in terms of impact, the minimal core pressure and the maximum curva-

ture of MSLP isolines associated with an identified track both have to be within the lowest respectively highest 5% of all tracks at least once within the defined radius. This constraint reduces the number of cyclone tracks significantly. However, it represents a necessary approach since a cyclone that is very intense somewhere over the Atlantic is unlikely to embody a severe damage potential for Europe. For

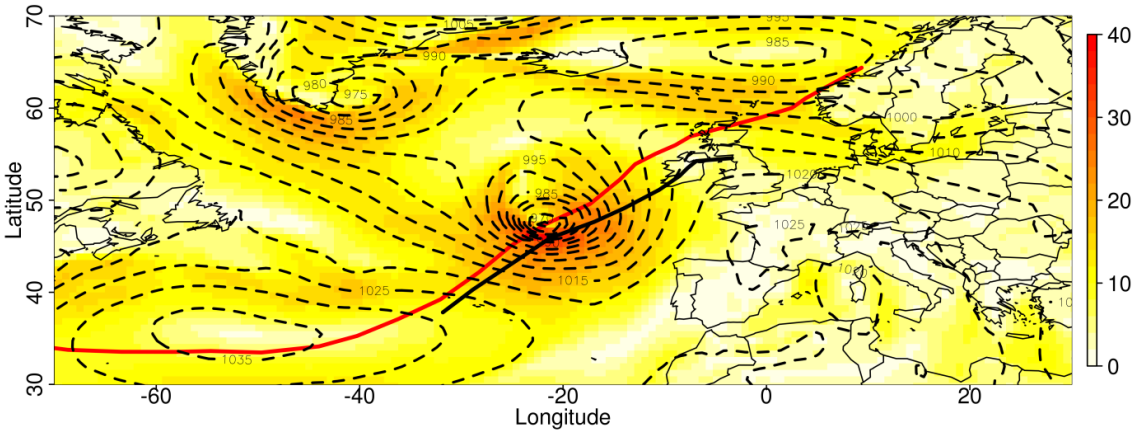


Figure 5.2.1: *Instantaneous wind speed of an arbitrary identified windstorm in System 4, overlayed by the MSLP field in black dashed lines. The red line represents the entire identified cyclone trajectory, the black line the entire associated windstorm track. As expected the area of maximum wind speeds is south of the core of the cyclone.*

illustration purposes of the tracking algorithm Figure 5.2.1 depicts a snapshot of the wind field as well as an identified cyclone track (red) together with its matched windstorm track (black). The matching of the two events was carried out according to the algorithm implemented by Nissen et al. (2010). The MSLP field is depicted as a black overlay. As expected the maximum wind speed can be found just south of the area of minimum core pressure of the cyclone. Evidently the cyclone trajectory is significantly longer than the windstorm track. This is due to the fact that a windstorm per definition is only tracked as long as the local 98th percentile of wind speeds is exceeded whereas the cyclone is tracked over its entire lifetime from

cyclogenesis to cyclolysis. In total more than 11,000 windstorms could be identified that fulfilled the area criterion for the British Isles amongst the 1581 virtual years of data which is an average of around 7 windstorms per December-March period. This compares to about 7,000 storms affecting Germany and the Benelux, around 8,000 for the Scandinavian region and approximately 15,000 affecting the Central European region. The exact numbers can be found in Table 1 in Chapter 5.4.1. Due to the constraint for the tracked cyclones the number of identified events is distinctively smaller. There are around 3,000 extreme cyclones affecting the British Isles, around 700 for Germany and the Benelux, 2300 for Scandinavia and roughly 1000 for Central Europe. Even though the numbers are smaller compared to the number of windstorms for every region, the sample size is still large enough to apply extreme value statistics. Clearly there is also an overlap between the different regions; however there are always windstorms/cyclones that are unique for each region. Additionally the overlap does not affect the estimation of the uncertainty within each particular region.

5.3 Methods

5.3.1 Clustering technique

In order to categorize the identified trajectories for cyclones and windstorms the regression mixture models clustering method proposed by Gaffney et al. (2007) was implemented in FORTRAN. This paper will only provide a short summary of the method, for a full description of the subject the reader is referred to their original study.

The initial position of every identified trajectory is subtracted from each respective pair of longitude and latitude coordinates in order not to cluster the tracks based on their origin but solely based on their shape. The initial location of a cyclone might influence the shape of the track, as we discard tracks, however, that never make landfall in Europe we can assume all of the tracks are shaped "regularly". The amount of resulting clusters has to be defined prior to the clustering. In agreement with Gaffney et al. (2007) three clusters are chosen to classify the windstorm/cyclone tracks. This choice is based on their results as well as on some qualitative inspection. Additionally three clusters provide a coarse overview as this study tries to understand the big picture of cyclone tracks, whereas with many clusters the clustering becomes more and more fuzzy. The trajectories are modelled via a second order polynomial which was determined to perform best via cross-validation and which is also in accordance to Gaffney et al. (2007).

The concept of the probabilistic clustering with regression mixture models is firstly to learn all the parameters (regression and covariance matrices) for all $K = 3$ clusters and in a second step to decide in which cluster a respective trajectory is most likely to be in. In other words, every trajectory is fitted by a quadratic polynomial and based on the coefficient matrices of this polynomial the algorithm calculates probability weights w_{ik} with $\sum_{k=1}^{3} w_{ik} = 1$. Subsequently these weights decide in which of the $K = 3$ clusters a respective track is most likely to be in. Starting from a random initial probability for every trajectory, three regression matrices β_k of size 3x2, 3 covariance matrices σ_k of size 2x2 and a probability w_{ik} is computed by the Expectation-Maximisation (EM) algorithm.

The EM algorithm is widely used for estimating the maximum-likelihood parameters in connection with regression mixture models (Dempster et al., 1977, McLachlan and Krishnan, 2007). It is a two-step algorithm as initially both of the parameters (β_k and σ_k) as well as the cluster assignment are unknown. The algorithm is iterated until the increase in the maximum-likelihood estimation falls below a certain threshold. A drawback of the EM algorithm is its potential to only find a local maximum in the maximum-likelihood surface. To increase the chance of finding a global maximum the algorithm is run 50 times with 50 different random starting weights. The events are hard-clustered by assigning a trajectory to the cluster featuring the largest of the three probability weights ("winning weight").

5.3.2 Analytical techniques – Windstorms

After assigning every windstorm trajectory to one of the three clusters the dynamical features of the associated events are examined. The intensity of a windstorm is given by the Storm Severity Index (SSI), an objective measure for the severity of a storm based on the cubic exceedance of the local 98th percentile of wind speeds (Leckebusch et al., 2008a). The SSI is part of the output of the windstorm identification and tracking algorithm. It is calculated and accumulated on a grid cell level for every time step of a respective windstorm. In order to assess the damage potential of a storm for a specific region, only SSI values within the defined radius around a region are added up. By applying means of extreme value statistics, i.e. fitting of a Generalised Pareto Distribution (GPD, e.g. Coles, 2001) to the excesses over a large threshold of SSI values (Peak over Threshold (POT) approach), return levels of windstorm intensities for every region and cluster can be estimated. The POT approach has been adapted by many other studies in connection with excessive

precipitation (e.g., Vrac and Naveau, 2007 or Cooley et al., 2007), wind speeds (Kunz et al., 2010) and also SSI values (Donat et al., 2011 or Held et al., 2013). The confidence intervals of the return levels are estimated via profile likelihood (Coles, 2001) as the intervals become highly asymmetric for upper bounds and high return periods. The parameter estimation of the GPD and the calculation of the confidence intervals are implemented in R with the help of the *ismev* library (Heffernan et al., 2012). In order to account for the probabilistic nature of the clustering, the cluster weights are used to weight features of the events associated with each trajectory. If, for example, the winning weight was 0.75 an SSI value of 1 will result in a weighted severity of 0.75. Given the shape parameter of the GPD distribution is negative; an expected maximum value of SSI values can be estimated. Additionally some general storm features like celerity, duration and average intensity are estimated for every cluster and region. MSLP anomaly composites are created as well in order to gain insight in the large scale conditions that predominate for the three clusters. This is done by averaging the anomaly MSLP fields of the windstorm days for the three clusters.

5.3.3 Analytical techniques – ETCs

The intensity measures of cyclones are denoted by the curvature of the MSLP field c and the minimal core pressure p of the cyclone. As it is difficult to relate curvature (especially due to its unit of $\frac{hPa}{deg.lat^2}$) to the extreme nature of an ETC the values are rescaled to an adapted Rossby number Ro. The Rossby number is a dimensionless estimation of the ratio of inertial force versus Coriolis force. U and L describe the characteristic velocity and length of a phenomenon, in this case of an ETC, and f represents the Coriolis parameter. With a characteristic length of 1,000,000 m, a

characteristic velocity between 10-100 m/s and a Coriolis parameter of $10^{-4}1/s$ the Rossby number of an average ETC is around $0.1 - 1$. The prominent definition of Ro is given in Equation 5.3.1a

$$Ro = \frac{U}{L * f} \tag{5.3.1a}$$

$$Ro* = \frac{c}{\rho * f^2} \tag{5.3.1b}$$

In order to rescale the curvature values they are scaled with the density of air and the square of the Coriolis parameter so that the result is dimensionless once again. With characteristic values of $f = 10^{-4}\frac{1}{s}$ and $\rho = 1\frac{kg}{m^3}$ and assuming 1°of latitude to be around 100,000 m the resulting values are actually of the same magnitude as the values for curvature. However, they become dimensionless and they can be related to an average Rossby number; thus more readily interpretable. As shown in Chapter 5.4.2 values of $Ro*=4$ are not uncommon for intense ETCs.

Both features still do not guarantee a potential high-impact storm event, however they both serve as very good proxies. As only cyclones are selected that range among the top 5% of all cyclones in terms of p and $Ro*$ within the defined areas around the regions, the chances are increased for the cyclone to have a severe impact in terms of wind speeds. Similar to the procedure applied to the windstorms events, a GPD is fit to the distribution of excesses of each $Ro*$ and p. That way return levels of minimum core pressure and $Ro*$ as well as their uncertainties can be estimated for every region and cluster. These maximum/minimum values can be compared and a cluster of the highest intensity potential can be determined for every region. Furthermore, a potential upper/lower limit of the $Ro*$/core pressure and their uncertainty can be

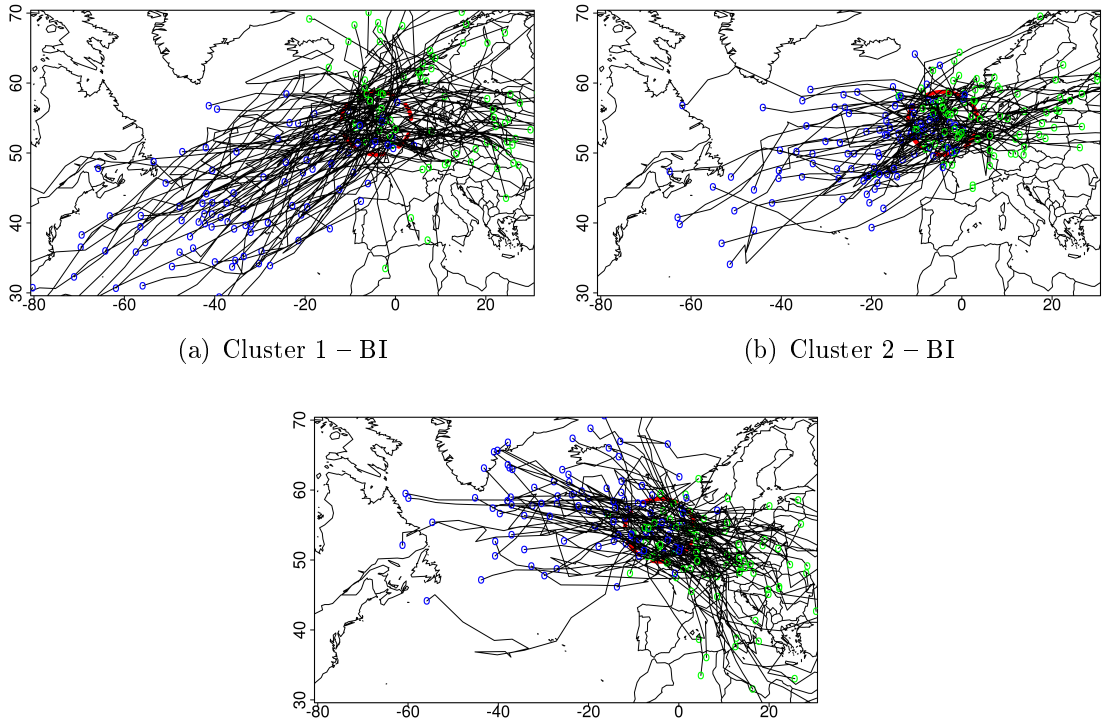
estimated. For a general overview of the identified clusters common attributes of cyclones like celerity, average core pressure and duration are examined and compared amongst the clusters and for the different regions.

5.4 Results

5.4.1 Windstorms

The three identified clusters for the windstorm trajectories can be seen in Figure 5.4.1 and 5.4.2. Figure 5.4.1 illustrates windstorms affecting the British Isles (BI), whereas Figure 5.4.2 depicts storms for Germany and the Benelux (GEBE) countries. The clusters for the other two regions look very similar (not shown). Clearly there are three general directions of progression of the tracks. The first cluster features events that follow the well-known North Atlantic storm track crossing the Atlantic in a north-easterly direction. For all the windstorm tracks it has to be kept in mind that the actual cyclone track might be a lot longer as windstorms by definition only exist whilst the 98^{th} percentile is exceeded. Cluster 2 includes events that are generally shorter than the ones in the first cluster: Their paths reflect a straighter West-East progression across the Atlantic compared to the first cluster. Events in the second cluster tend to be detected later, thus more towards the Central Atlantic. From there they proceed in an almost straight track towards the East. This cluster also contains some events that are only detected when the system is already close or even within the defined radii, i.e. the British Isles. Events that are included in Cluster 3 look distinctively different compared to the other two identified clusters. Events featured in cluster three are identified over the Denmark Straight usually between Greenland and Iceland or just south of Iceland. The

tracks then follow a south-easterly trajectory, approaching Europe by crossing the Northern Sea. This is also the cluster that differs most from the clusters identified by Gaffney et al. (2007). Their clusters D and H resemble Cluster 1 and 2 of this study whereas their cluster V cannot be identified for the tracks of System 4 at all. The progression of the tracks within these clusters is almost reversed: south-north



(a) Cluster 1 – BI

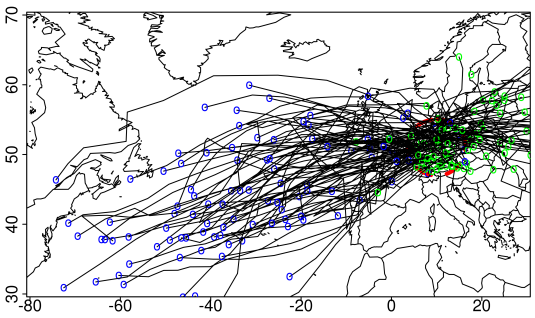
(b) Cluster 2 – BI

(c) Cluster 3 – BI

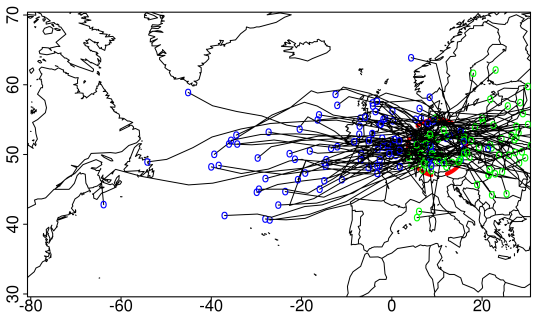
Figure 5.4.1: *The three clusters that were identified by the probabilistic clustering method for windstorms affecting the British Isles. Cluster 1 (top left) with its southwest to northeast progression, Cluster 2 (top right, west to east) and Cluster 3 (bottom, northwest to southeast). 100 random tracks are shown for each of the clusters. The blue points mark the beginning of the track whereas the green ones depict the last time step of each trajectory. The red circle encloses the British Isles and has to be crossed at least once in the life cycle of the windstorm.*

compared to northwest-southeast. This is certainly due to the fact that for the present study only events were considered that actually affected one of the defined

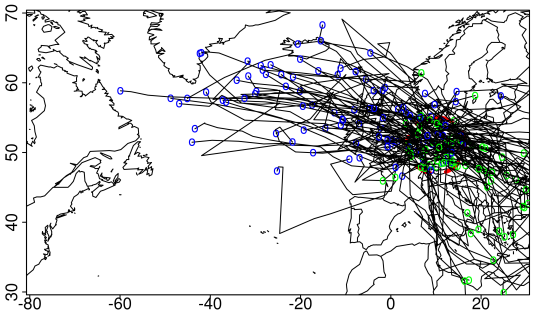
regions, whereas Gaffney et al. (2007) clustered all tracks that could be identified for their entire domain (i.e. have their origin further west). The number of storms



(a) Cluster 1 – GEBE



(b) Cluster 2 – GEBE



(c) Cluster 3 – GEBE

Figure 5.4.2: *Same as above except for windstorms affecting Germany and the Benelux (GEBE).*

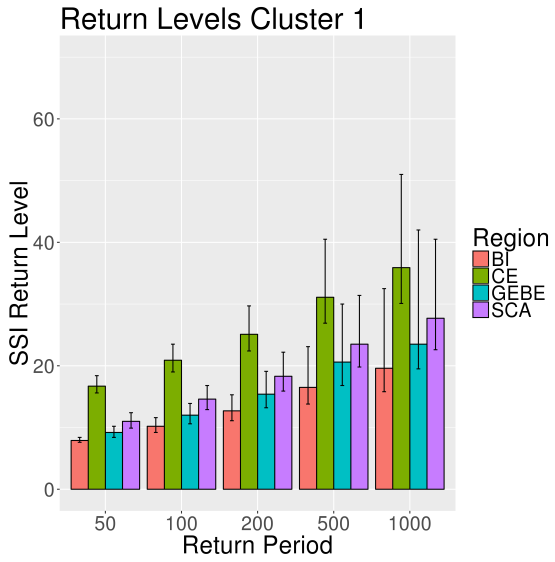
for every region and cluster can be found in Table 1. Clearly the Cluster 2 is the most frequent across all 4 regions as it makes up around 50% of all storms except for the Scandinavian region where it accounts only for about 40% of all the storms. Cluster 1 and 2 feature similar numbers for 4 all regions with Cluster 1 being more frequent in the BI and GEBE whereas Cluster 3 is more frequent in Central Europe (CE) and Scandinavia (SC).

Figure 5.4.3 provides (weighted) SSI return level estimates and their uncertainties for every region and cluster. Overall Cluster 1 represents the cluster with the most

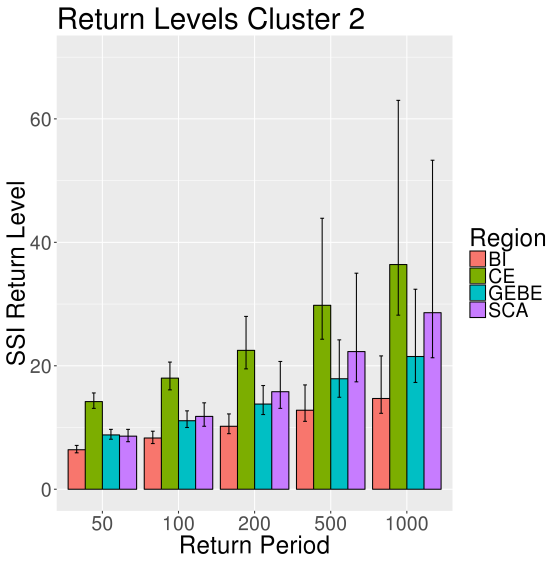
Table 5.4.1: *Number of windstorm tracks for all of the three clusters for the four different regions studied. The arrow behind the clusters represents the general path of the trajectory for each cluster.*

Region	*Cluster 1* ↗	*Cluster 2* ⟶	*Cluster 3* ↘	*Total*
British Isles	3182	5374	2482	11038
Germany and Benelux	1599	3911	1512	7022
Central Europe	3471	7244	4182	14897
Scandinavia	2318	3220	2863	8401

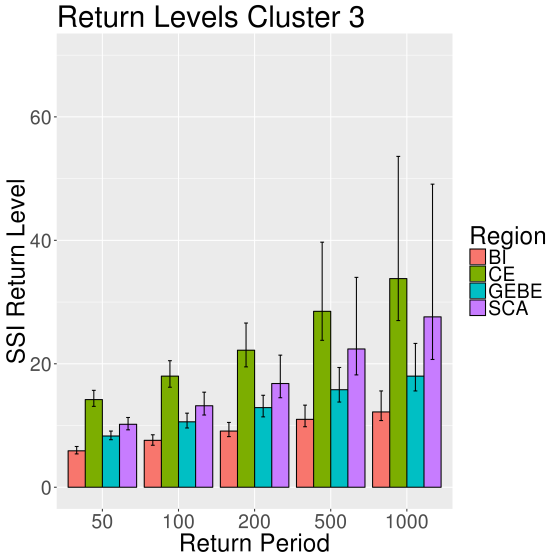
intense storms especially for the lower return periods. Particularly for the BI and GEBE the SSI return levels for Cluster one are the largest values for all periods, making it the most hazardous of the 3 clusters in terms of potential damage for these two regions. In the same way Cluster 3 appears as the least hazardous cluster as it contains the lowest SSI return levels across all return periods. The difference in magnitude between the clusters can be quite substantial as for example a 100 year event for the British Isles within Cluster 1 would be a 200 year event for Cluster 2 or almost a 500 year event for Cluster 3. Clearly the uncertainty of Cluster 3 is the lowest amongst the three clusters for BI and GEBE. Even intervals for the very large return periods appear almost symmetric around the estimated value whereas the confidence intervals for the high return levels of Cluster 1 and 2 are highly asymmetric towards the larger values. This reflects the larger uncertainty towards very high impact windstorm events. For CE and SC the order of the most hazardous cluster appears slightly different. For lower return periods Cluster 1 still represents the most hazardous cluster, however for the 500-year and 1000-year return level Cluster 2 for CE and Cluster 3 for SC emerges as the most hazardous one, especially when considering the confidence intervals. For the 50-, 100- and 200-year return period, Cluster 2 and 3 are virtually identical for CE.



(a) Cluster 1



(b) Cluster 2



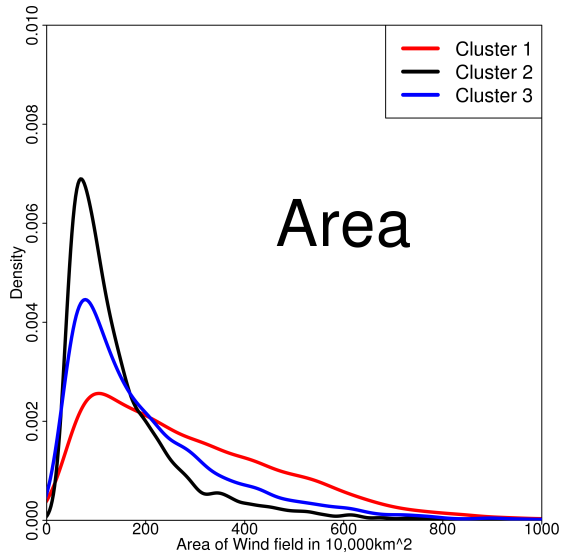
(c) Cluster 3

Figure 5.4.3: *Weighted return levels for given return periods of SSI values for all four regions. The 95% confidence intervals are marked with whiskers and are calculated via the profile likelihood method.*

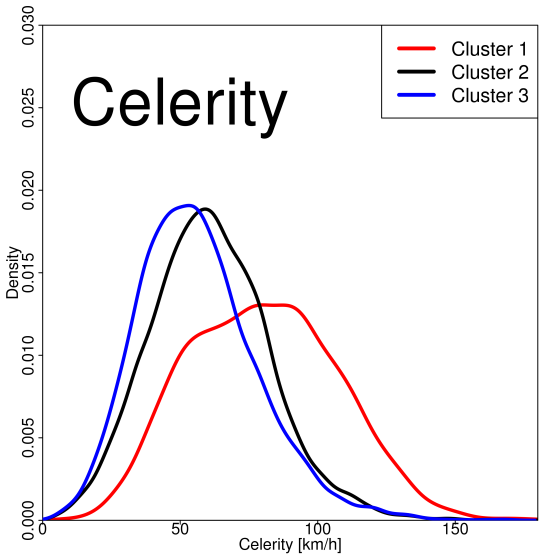
Interestingly Cluster 3 for SC is potentially more hazardous for return levels including the 500-year period than Cluster 2 which, in contrast to all other regions, represents the least hazardous cluster for SC. Overall, return levels for the 1000-year

period for CE and SC look very similar across all three clusters. Opposed to the BI and GEBE, Cluster 1 represents the cluster with the least uncertainty whereas Cluster 2 appears as the most uncertain, particularly for Central Europe where the upper confidence interval for the 1000-year return level reaches up to a value of 63. This is due to extremely intense outliers within Cluster 2 and also Cluster 3. Owing to the amount of storms per cluster and region however, the confidence intervals are still comparatively small considering the very large nature of the return periods.

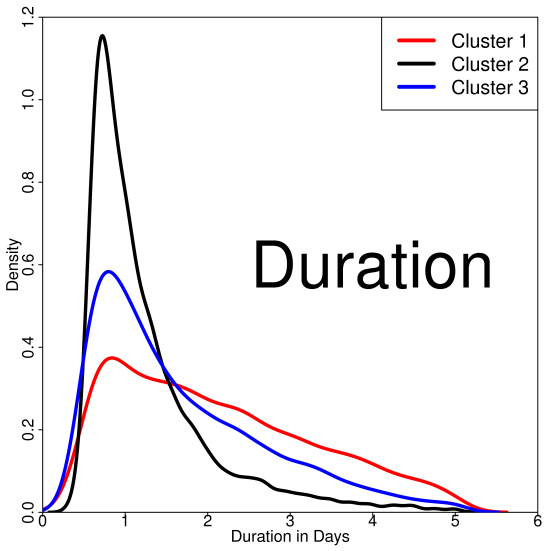
Generally SSI values over the mainland of Europe are systematically larger than over the BI and parts of Scandinavia. This is a result of the construction of the SSI: As the definition of the SSI does not take into account the shape of the distribution of wind speeds past the 98th percentile, SSI values are systematically higher for grid cells in which windstorms occurrences are less frequent (Walz et al., 2017). From an impact perspective this assumption is valid as it can be expected that the infrastructure in these areas might not be as adapted to frequent windstorm events as it is in areas within the main storm corridor (i.e. South of France vs. the British Isles). Figure 5.4.4 provides the weighted windstorm features for BI. Clearly Cluster 1 is the most extreme cluster in all three features addressed. Particularly the celerity of the events in Cluster 1 is considerably larger than for the other two clusters which appear very similar. Cluster 2 features by far the shortest events. This confirms the impression that could already be drawn from the trajectories in Figure 5.4.1. The larger area of windstorms in Cluster 1 partly explains the increased severity of windstorms as the numbers examined in Tables 2 and 3 are integrated values over all grid cells affected within the radius around a particular region. Interestingly Cluster 1 also entails the largest events in terms of area for CE and SC. As the largest



(a) Area



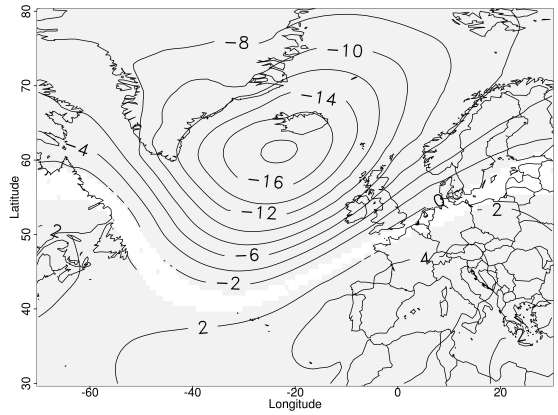
(b) Celerity



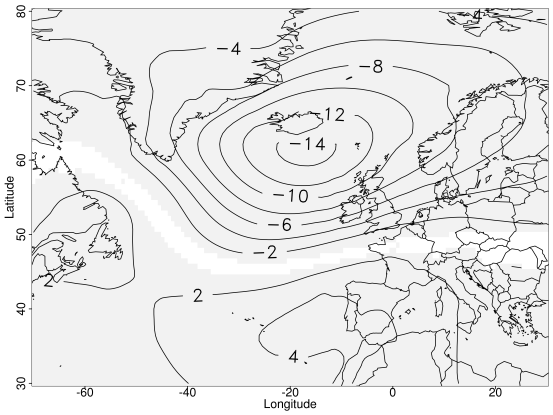
(c) Duration

Figure 5.4.4: *Weighted windstorm features found for the three defined clusters for the British Isles region. Area of the tracked windstorm in units of 10,000km2 (left), average celerity of the storm in km/h (centre) and duration of the windstorm events per cluster in days (right).*

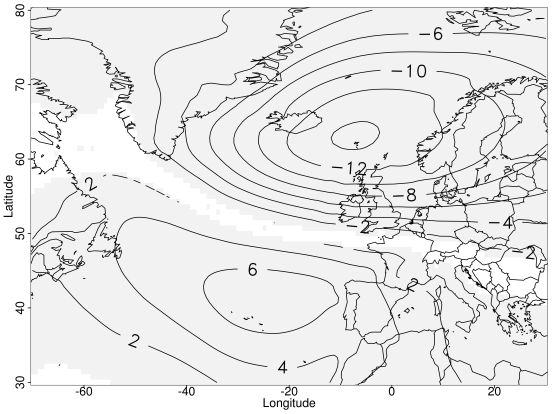
return level for the 1000-year event is found for Cluster 2 however, this means that there is no linear relationship between area and intensity of a windstorm event. The



(a) Cluster 1



(b) Cluster 2



(c) Cluster 3

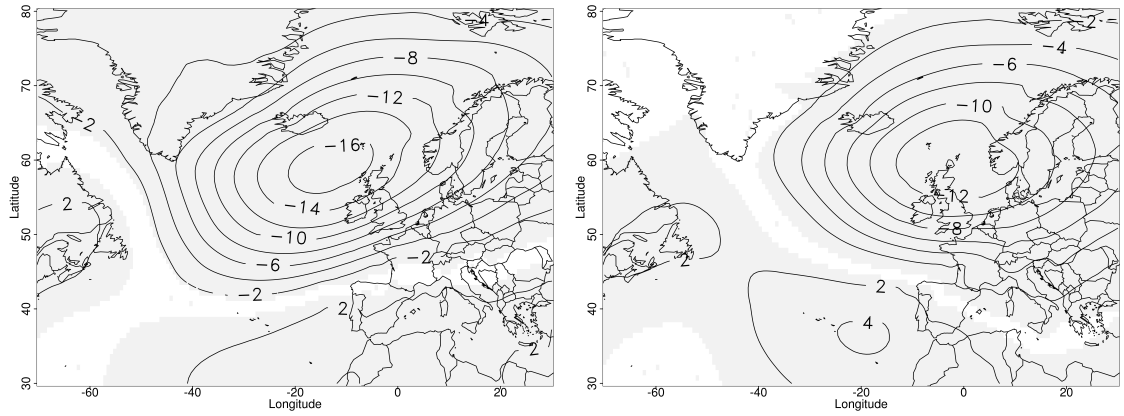
Figure 5.4.5: *MSLP Composites of windstorm days for each of the three clusters for storms affecting the British Isles. Cluster 1 (top left), Cluster 2 (top right) and Cluster 3 (bottom). All grey shaded areas are significant at the 99% interval*

meteorological conditions that are predominant during windstorm events for each cluster are presented in Figure 5.4.5. The panel shows MSLP anomaly composites of all the days on which windstorms were identified for each cluster for the BI region. All three figures show the stationary low pressure system that is associated with the Icelandic Low. The shape and location of this system, however, is different for all three clusters. Whereas for Cluster 1 and 2 the centre of the MSLP composite lies southwest of Iceland, the centre for Cluster 3 is shifted towards the northeast. This

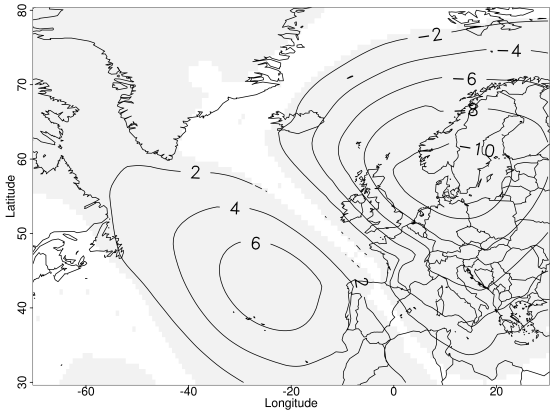
explains the different trajectories for the major part of storms in Cluster 3 compared to the other two clusters as the MSLP gradient points in a different direction.

The direction of the MSLP gradient for each cluster is approximately equivalent to the main path of the trajectories within each cluster, thus the direction of the large scale atmospheric flow under the geostrophic wind approximation. Windstorms in Cluster 3 appear upstream ("backside of the wave") of the atmospheric wave, whereas windstorms in the other two clusters are the result of a more downstream development. Compared to the quiescent composites (not shown) the gradient is considerably stronger explaining the higher potential for storminess for each of the three identified clusters.

The MSLP patterns for the windstorms affecting the CE region are presented in Figure 5.4.6. The overall MSLP anomalies look similar to what was shown for the BI region, however the minimum of the anomalies is shifted further to the East, for Cluster 1 in particular. In accordance to the findings for the cyclones (see Chapter 5.4.2), events in Cluster 1 are lower in core pressure compared to the other two clusters (-18 hPa vs. -14 hPa; compare the following chapter). The MSLP composites for the other 3 regions are according to the one shown for the BI and CE. The associated MSLP anomalies for the windstorms affecting the CE region draw a similar picture, even though the Icelandic Low is shifted more towards the South. In view of the more southern region of CE this is according to the expectation. The composite for Cluster 3 emphasises the shifted low pressure system even more as the core of the system is located over southern Scandinavia which enables windstorms to travel in a south-easterly direction upstream of the associated low pressure system.



(a) Cluster 1 (b) Cluster 2



(c) Cluster 3

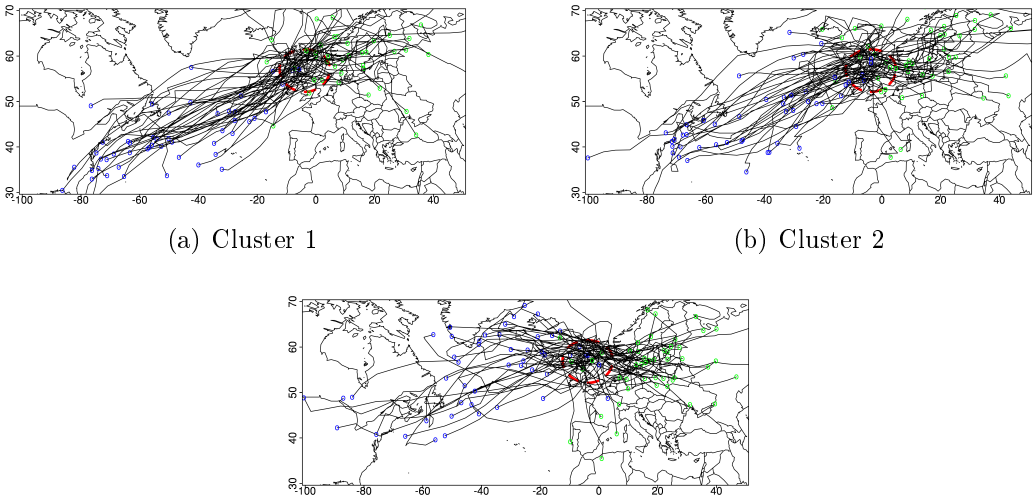
Figure 5.4.6: *Same as in Figure 5.4.5 but for Central Europe.*

5.4.2 ETCs

The 3 identified clusters for the cyclone trajectories only differ marginally from the clusters found for the windstorms and are presented in Figure 5.4.7. However by classifying the cyclones in three different clusters we can give estimates of a *minimum* (*maximum*) core pressure (Rossby number Ro) for different return periods based on a cluster. Depending on the shape parameter of the GPD fit we can even estimate an absolute to be expected minimum (maximum) of core pressure (Ro). This way a we can estimate both the most extreme cluster and the most uncertain

cluster with regards to extremity.

Whereas Cluster 1 looks very similar, the trajectories in Cluster 2 and 3 look slightly different to the ones for the windstorms. As windstorms are only identified and tracked if the local 98th percentile of wind speeds is exceeded, they tend to be considerably shorter than cyclone tracks. Thus, they are usually identified at a later stage of the development of a cyclone. That is the reason why many of the cyclone tracks in Cluster 3 appear much longer than the equivalent trajectories for windstorms. Most of the windstorms in Cluster 3 have their origin in the North Atlantic and the Labrador Strait. The origin of the associated cyclone, however,



(a) Cluster 1

(b) Cluster 2

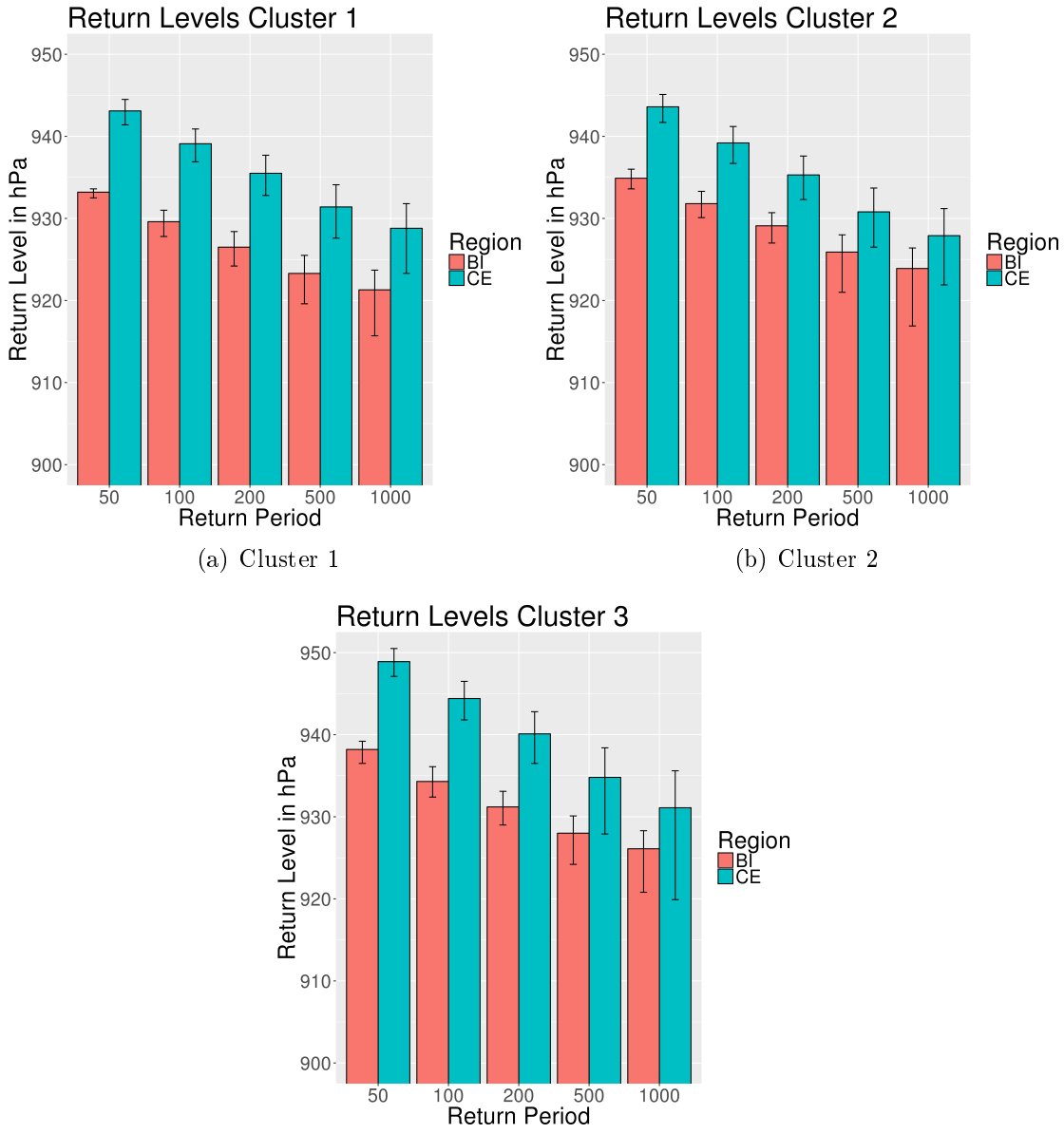
(c) Cluster 3

Figure 5.4.7: *The three clusters that were identified by the clustering probabilistic technique for intense cyclones affecting the British Isles. Cluster 1 (top left) with its southwest to northeast progression, Cluster 2 (top right, west to east) and Cluster 3 (bottom, northwest to southeast). 50 random tracks are shown for each of the clusters. The blue points mark the beginning of the track whereas the green ones depict the last time step of each trajectory. The red circle encloses the British Isles and has to be crossed at least once in the life cycle of the ETC.*

is further upstream off the coast of Newfoundland. Hence, most cyclone tracks in

Cluster 3 travel in a north-easterly direction whilst they intensify until they reach the Labrador Strait. This is the point where they will be detected by the windstorm tracking algorithm. Subsequently they travel in a south-easterly direction (upstream of the wave) which leads them towards the European mainland. Similarly, cyclones in cluster 2 are identified long before the windstorm tracking algorithm associates the respective windstorm to the cyclone. In this case they are only identified as windstorms halfway across the Atlantic Ocean which lets them appear so much shorter compared to the entire cyclone track. Generally, it is harder to distinguish between cyclone Cluster 1 and 2 both qualitatively (i.e. Figure 5.4.7) and quantitatively as demonstrated by Figure 5.4.8 and 5.4.9. Figure 5.4.8 provides return levels for the same return periods as previously for minimum core pressure of a cyclone within the defined area for BI and CE. The return levels for Cluster 1 and 2 look very similar for both regions, especially for CE where they are virtually the same throughout all return periods. In accordance with the intensities of windstorms, Cluster 3 features the potentially least strong events, revealed by the largest minimum core pressure. Table 5.4.2 provides an estimate of the lowest potential minimum cyclone core pressure to be expected for each cluster. This estimation is based on the (negative) shape and scale parameter of the respective GPD that was fit for each cluster. As before Cluster 1 and 2 feature virtually the same value whereas Cluster 3 exhibits a larger, thus less intense, core pressure.

It is not possible to calculate a lowest potential core pressure for Cluster 3 for CE as the shape parameter of this particular GPD is positive which means that no upper limit can be estimated. Due to the abundance of cyclone events, the uncertainty in estimating the return values is fairly low considering the scarcity of these events.



(a) Cluster 1

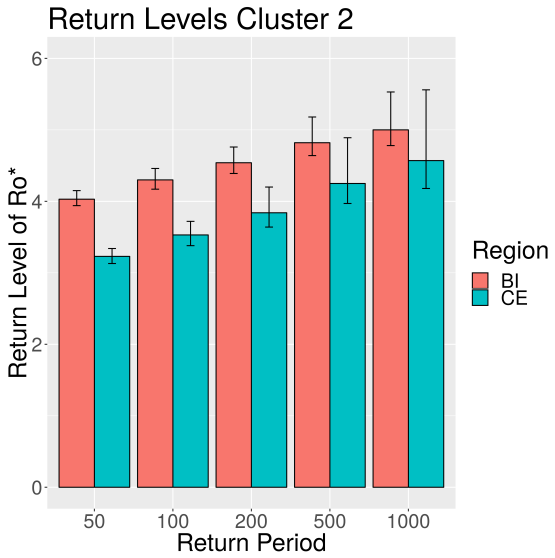
(b) Cluster 2

(c) Cluster 3

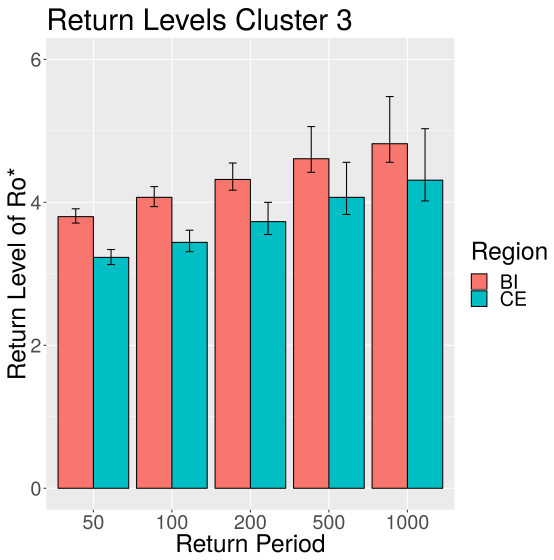
Figure 5.4.8: *Weighted return levels of minimum core pressure for given return periods of the British Isles and Central Europe. The 95% confidence intervals are calculated via the profile likelihood method and are presented in brackets.*

The confidence interval for a 500-year return level for BI and Cluster 1 for example only ranges from 919-925 hPa which is a fairly accurate estimation. Return levels of the adapted Rossby number $Ro*$ are presented in Figure 5.4.9. $Ro*$ is a number

(a) Cluster 1



(b) Cluster 2



(c) Cluster 3

Figure 5.4.9: *Weighted return levels of the Rossby number Ro* for given return periods of the British Isles and Central Europe. The 95% confidence intervals are calculated via the profile likelihood method and are presented in brackets.*

to describe the ratio of geostrophic relative vorticity to Coriolis parameter. An average ETC features values of Ro^* of around 0.1–1. Thus the presented values indeed represent intense cyclonic activity. Generally these values are in accordance

with the results for the minimum core pressure and also the intensity of windstorms. Events in Cluster 1 feature the highest *Ro** return levels, whereas the ones in Cluster 3 appear as the lowest. Compared to the core pressure values however, there is a larger difference between Cluster 1 and Cluster 2. Table 5.4.2 also features estimates of the expected upper limit of *Ro**. Interestingly Cluster 3 features the largest upper limits even though the return levels were lower than both of the other clusters. One of the reasons for that could be the larger uncertainty of Cluster 3 compared to the other two clusters. Especially the uncertainty in Cluster 1 is remarkably low. Even for a 1000-year event the confidence interval comprises the estimate really closely. Additionally the confidence interval is only slightly asymmetric, whereas

Table 5.4.2: *Potential minimum core pressure (in hPa) and Ro* (dimensionless) for all three clusters for the BI and GEBE region*

Region	*Cluster 1* ↗	*Cluster 2* ⟶	*Cluster 3* ↘
British Isles	911 | 6.2	909 | 6.4	918 | 7.3
Germany and Benelux	909 | 5.8	907 | NA	NA | 7.3

the confidence intervals for the 1000-year windstorm intensities are considerably asymmetric (e.g. Figure 5.4.3). In terms of intensity the events that affect the BI are clearly stronger compared to the ones affecting CE. This is in contrast to the intensities of the windstorm events as the largest potential SSI values are estimated for CE and the lowest for the BI. As mentioned before, this is in line with the findings from Walz et al. (2017). Furthermore there is not necessarily a linear relationship between a large curvature/*Ro**, a small core pressure and high surface winds. Some of the windstorms might not be extreme in terms of pressure and curvature/*Ro** and vice versa.

5.5 Summary and Discussion

The probabilistic clustering technique proposed by Gaffney et al. (2007) was implemented for the purpose of investigating the statistical uncertainty of extreme extratropical cyclones. Two different identification and tracking algorithms for identifying extreme cyclone/windstorm events were applied: one based on exceedances of the local 98th percentile of wind speeds (Leckebusch et al., 2008a), the other one based on maxima in MSLP curvature (Murray and Simmonds, 1991). The events were identified in retrospective forecasts of the ECMWF Seasonal Forecast System 4 that comprises 51 ensemble members. The 51 members are used in a climate archive context similar to Osinski et al. (2016), thus every member is treated as an individual artificial reality of 31 years resulting in more than 1500 years of physically consistent data. As the forecast skill for very high local wind speeds is relatively small (Walz et al. (2018b)), the members can be seen as statistically independent since statistical dependence and forecast skill can be seen as equivalent. The clustering was implemented in accordance with Gaffney et al. (2007) who used three clusters and quadratic polynomials to model the individual trajectories. Cluster 1 and 2 are equivalent to the clusters identified in their paper. The trajectories assigned to Cluster 3, however, are entirely different as it features windstorm events that approach Europe in a south-easterly direction whereas their third cluster mainly features events of south-north displacement. On the one hand this could be due to the different tracking algorithm (wind based vs. pressure based). As mentioned above, windstorm tracks are only about half as long as cyclone tracks. This affects their shape and thereby the coefficients of the polynomial that is used to fit them. On the other hand the different cluster might be due to the geographical selection process of the windstorms used for the clustering.

Gaffney et al. (2007) use identified cyclones from the entire North Atlantic area while this study only considers trajectories that cross a certain defined area, i.e. the British Isles. Thus, windstorms that do not affect any of the defined regions and that potentially travel in a south to north direction are generally excluded from the clustering algorithm. The statistical analysis was carried out via a GPD model which is part of the theory of extreme value statistics (Coles, 2001). Return periods of windstorm intensities (quantified by the SSI) for all three clusters and all four defined regions were determined. Cluster 1, which contains storms that cross the Atlantic diagonally following the classic Atlantic storm track, appears as the cluster that includes the potentially most severe windstorm events, in particular for BI and GEBE. Due to some very intense events included in Cluster 1 the uncertainty of the cluster for the very large return periods (>=500 years) is considerably larger than for the other two clusters. Considering the very large return periods, however the overall absolute uncertainty for BI and GEBE is fairly small. Especially the confidence intervals for Cluster 3, which contains the potentially least severe windstorms, are reasonably small and also symmetrical. These exact intervals can be achieved by utilizing more than 11,000 windstorms for the BI and more than 7,000 for GEBE.

The intensities for CE and SC are similar; however the 1000-year event for Cluster 2 is larger than the equivalent value for Cluster 1. In the same way the confidence intervals are larger and less symmetrical compared to the ones for the BI and GEBE. Both of this is due to some outliers in intensity being part of Cluster 2. Additionally, the differences in intensities between the clusters are generally lower, especially for SC where the 1000-year event is essentially the same across all three clusters. The

intensity values are generally difficult to compare across the four regions as the SSI is dependent on the local wind speed climatology Walz et al., 2017). Large SSI return values for CE or GEBE can, however, be interpreted as a lack of preparedness of the local infrastructure in some of the countries included in the region against these high-impact storms due to their infrequent occurrence. Despite these discrepancies between the clusters, the return levels provide a valuable quantification of the severity and especially their uncertainty for the four different regions and three different clusters. In terms of windstorm features Cluster 1 contains the largest and also longest events. As the intensity is an aggregated quantity this would explain the more severe events in this cluster. However, even though Cluster 2 contains the smallest and shortest events also for CE and SC their intensity is higher for those regions. Thus there is not a clear linear relationship between area, duration and intensity. Regarding the celerity of the events, there is a large difference between Cluster 1 and the Clusters 2 and 3. Events in Cluster 1 travel considerably faster than the ones in the other two clusters.

The MSLP anomaly composites for the three different clusters indicate a successive eastward displacement of the Icelandic Low for Cluster 1, 2 and 3 respectively which allows for the cyclones and the associated windstorms to develop more upstream, thus pursuing a south-eastward track compared to a northward movement. Compared to composites of days without any storm activity (quiescent cluster), the MSLP gradient over the North Atlantic is substantially stronger resulting in more intense (i.e. deeper core pressure) cyclone systems that get eventually identified as a windstorm. The associated cyclones for windstorms in Cluster 1 are distinctively lower compared to the other two clusters. This is in good agreement with the

findings for the tracked cyclones as well. Compared to the composites created in Gaffney et al. (2007) no negative NAO conditions are evident for Cluster 2 (their H–Cluster). This, however, is most likely due to the fact that windstorms hardly occur during negative NAO conditions (Donat et al., 2010). The Clusters 1, 2 and 3 link to the Primary Storm Cluster (PSC) 1, 2 and 4 in Leckebusch et al. (2008b) respectively. The PSC 3 of their study can be considered as a hybrid of Cluster 1 and 2. They assigned more than 70% of the historical extreme storms to these four clusters. This in turn means that the clusters identified in this study proof to be a valid characterization of the spatial variability of extreme windstorms.

The three identified clusters for the cyclone tracks are less distinct compared to the windstorm clusters. Qualitatively it is more difficult to find a difference between Cluster 1 and 2 as the trajectories of both clusters are fairly similar. However, similar to Cluster 2 for the windstorms, the trajectories in Cluster 2 for the cyclones tend to be shorter and pursue a more direct eastward track compared to the very distinct diagonal track of Cluster 1. Cluster 3 of the cyclones is comparable to Cluster 3 of the windstorms as the general progression of the identified tracks is also from the central North Atlantic towards the southeast. In contrast to the windstorm in Cluster 3 however, their origin is further west as windstorms are identified later in the life cycle of the associated cyclone. For that reason the cyclone trajectories in Cluster 3 follow are more "arc-like" path compared to the diagonal displacement of windstorms. The intensity of the cyclones is represented by the minimum core pressure and $Ro*$. As the cyclones identified by the tracking algorithm are not necessarily extreme per se, cyclones have to be amongst the most extreme 5% in both variables at least once within the defined areas around the four regions. Even

though that does not guarantee a high-impact storm it does serve as a good proxy for an extreme event.

This constraint reduced the amount of cyclones for every region drastically so that roughly around 500-1000 cyclones per region were considered. Bearing in mind the scarcity of these extreme cyclones however, the sample size was still large enough to estimate return levels for very large return periods accurately. In accordance with the return levels of the windstorm events, the lowest core pressure can be assumed to occur in a cyclone from Cluster 1. Even though there are fewer events per cluster compared to the windstorm clustering, the uncertainty is considerably smaller ($\pm 1\%$) for both the minimum core pressure and the Ro^*. Additionally the confidence intervals are fairly symmetrical which in turn is an indication for well-estimated confidence intervals. The reason for this good estimate is the lack of extreme outliers that made it easier to find a threshold and fit a GPD to the distribution of excesses. Economou et al. (2014) estimated core pressure extremes based on a statistical model. Their estimate for a 50–year event for the UK is approximately between 940–970 hPa (for a strongly positive NAO phase). This is higher than the estimates we could give based on System 4 (around 932-937 hPa depending on the cluster). However "only" 31 years of observational data were used for their study, so a reason for the higher estimates (and larger uncertainties) might just be the almost factor 50 times smaller original sample size.

The lowest ever recorded MSLP in the British Isles was 926 hPa in January 1884 (MetOffice, 2016) which would, depending on the cluster, represent a 500-1000–year event based on our analysis. This shows that the estimations made from System 4

lie within a physical possible horizon. The lowest potential core pressure of a cyclone is estimated to be around 910 hPa for both the BI and CE. This is more than 10 hPa below the 1000-year return level (and thus the lowest ever recorded value) so it can be assumed that an event of this magnitude is highly unlikely and should be regarded as a physical barrier rather than an actual event. Interestingly the largest potential $Ro*$ for both regions is found for Cluster 3 even though the return levels are the smallest of the three clusters. However, this is in accordance with the generally higher uncertainty of Cluster 3, expressed by the larger confidence intervals for both $Ro*$ and core pressure.

Overall it has to be kept in mind however that the data used here is a forecast product which, like any other data set, is not devoid of biases. Although a clear wind speed or MSLP bias has not been found in System 4 some studies (e.g. Torralba et al., 2017) suggest that a bias correction is necessary as the raw wind speeds of System 4 show too little variability in the ensemble mean. As shown in Chapter 4, however the entire ensemble spread of wind speeds capture the observations well. As this study is investigating individual members the bias in the ensemble mean is of small relevance.

The clustering approach represents a useful instrument to classify these rare extreme events and to determine large-scale differences between the different clusters. Due to the approach of considering the 51 ensembles of the retrospective seasonal forecast of the ECMWF System 4 as a vast statistical base, the uncertainty in intensity and extreme core pressure even for very large return periods could be estimated fairly accurate. For most regions an empirical physical barrier of core pressure and $Ro*$

could be estimated as well. As these barriers are well below/above the 1000-year return level, they act as an estimate of what could possibly happen rather than as a magnitude that is likely to occur. Windstorms/cyclones in Cluster 1 pose the largest threat in terms of potential damage for most of Europe as it features the largest return levels for all metrics addressed: Windstorm events in Cluster 1 are of the biggest size and of the highest celerity, thus making it the most hazardous cluster of the three.

Chapter 6

Impact of winter windstorms

"You don't need to pray to God any more when there are storms in the sky, but you do have to be insured."

– Bertholt Brecht, German poet and playwright in *The Mother, 1930*

Abstract

In this paper we investigate the feasibility and added value of using the seasonal hindcasts of the ECMWF System 4 as a hazard event set for European winter windstorms loss calculations. The windstorms are identified for every ensemble member and every year by an objective windstorm tracking algorithm. The loss is calculated directly from the obtained wind footprints via the open source natural catastrophe damage model climada for Germany, the UK, France and Spain and compared to the loss from ERA-Interim. The results show that the ensembles of losses in System 4 nicely capture the inter-annual loss variability of the reanalysis. Due to more than 1500 years of 'virtual reality' windstorm data from the hindcasts the return levels of extreme losses can be estimated fairly accurately. Based on System 4, the losses in the scale of 1990 (January, February, March and December including the prominent windstorm Daria) represent a 20-year event in Germany whereas they represent a 100-year event for the UK. Thus, a considerably shorter return period compared to return periods calculated from ERA-Interim alone. Further we investigate the link between the annual losses and large-scale drivers derived from mean-sea-level-pressure (MSLP) data in System 4. We can show that within System 4 there is a significant link between increased loss potentials for strongly positive North Atlantic Oscillation (NAO) phases for Germany and the UK as well as a reduced loss potential for Spain. The link between the other analysed indices is weak bar the East Atlantic (EA) pattern index. Thus, if the NAO in System 4 is correct we can assume that the windstorms in System 4 are useable. If this premise is given our study shows that the loss estimates and ultimately the return levels of losses from System 4 can be used in an operational way.

The following chapter is an edited version of the article currently under review:
Walz, M. A. & Leckebusch, G. C. (under review, submitted 13 September 2018). Windstorm-induced losses based on a hazard event set created from ECMWF System 4. Atmospheric Science Letters

This study was designed by Michael Walz and Gregor Leckebusch. All data processing and analysis was carried out by Michael Walz. The *climada* software used in this study was provided by David Bresch (`https://github.com/davidnbresch/climada`). The manuscript was written by Michael Walz. Gregor Leckebusch gave valuable comments which helped to bring out the added value of this study more clearly. This study is currently (March 9, 2019) under review.

6.1 Introduction and Motivation

Winter windstorms are extreme events that lead to considerable losses across Europe. Due to the large year-to-year variability in number and intensity of these storms the observational record of these high impact events is fairly small. Recent reanalysis projects like ERA-20C (Poli et al., 2016) cover a period of around a hundred years, however given the extremity of certain events (e.g. Daria 23-29 Jan 1990) the tail of the loss distribution still only features a handful of extreme losses associated with windstorms. As recent studies have shown there is also a spurious trend in this reanalysis data set (Bloomfield et al., 2018 or Befort et al., 2016). The lack of observations is often tackled by producing probabilistic event sets based on alteration of observed events (e.g. Schwierz et al., 2010). The way stochastic event sets are generated, however, does not necessarily require or account for any physical consistency within an event as it evolves through time. Thus, a key part of the loss modelling, the uncertainty in the hazard, is not adequately understood. This paper aims to investigate the feasibility of generating a set of physical consistent events to assess the related uncertainties in potential damage and loss.

The fundamental idea is to identify windstorms in the 51 members of the ECMWF System 4 seasonal forecast system (Molteni et al., 2011) and treat each member as a physical consistent realisation of a potential reality. This approach is similar to Osinski et al. (2016) who used the ECMWF EPS model to build a windstorm 'hazard set'. This will lead to a substantial increase in the available physically consistent sample of extreme events. The annual losses for four different European countries for every member of the ensemble are estimated from the tracked windstorm events with the help of the open source natural catastrophe damage model climada (Bresch

and Mueller, 2017).

Various previous studies have proposed the effect of large-scale drivers onto the intensity/frequency of cyclones (Pinto et al., 2009) and windstorms (Donat et al., 2010 or Walz et al., 2018a). In order to see whether this link is also represented within the seasonal forecast the estimated regional annual losses are set into context with model-internal large-scale driver time series (e.g. North Atlantic Oscillation; NAO; Hurrel, 1995).

Chapter 6.2 will provide an overview of the methods and data used for this study. It will also include a short description of climada. For a detailed description of the tool the reader is kindly referred to the official manual (climada manual; Bresch and Mueller, 2017). Chapter 6.3.1 will present the results of the loss estimation for the four countries including the large-scale driver composite analysis. Chapter 6.4 provides a summary and discussion as well as a brief outlook on future research.

6.2 Data and Methods

The hazard event set is based on the ECMWF System 4 hindcasts entailing the years 1982-2014 (Molteni et al., 2011). There are 51 ensemble members which are initialised every November 1st. In order to exclude any potential "real" storms in November from the analysis, only the months December until March are included in the hazard event set; as we can assume that by December the effect of the initialisation has vanished. The events are identified for 6-hourly 10m wind speeds within every single member using the WiTRACK algorithm (Leckebusch et al., 2008a, Kruschke, 2015 or Befort et al., 2016). By using a wind-speed-based tracking algorithm

we directly obtain the extreme wind field which can then be used for loss estimation in climada. This is a huge advantage compared to mean-sea-level-pressure (MSLP)-based cyclone tracking algorithms. In total, this results in more than 1500 years of 'alternative reality' storms (32 years x 51 members).

In order to set the loss estimated from System 4 in context with observations, windstorms are also tracked for the same years in ERA-Interim. The resolution of both the hindcasts and the reanalysis is T255 so that there is no systematic bias due to differences in model resolution. Large-scale driver time series are computed via an Empirical Orthogonal Function analysis (EOF) using 6-hoursly MSLP again for both System 4 and ERA-Interim. The loss calculation is done for the UK, Germany, France and Spain, thus the countries generating the most loss caused by winter windstorms. In accordance with the actuarial industry the loss is calculated for the entire year, thus losses for one year consist of January, February, March and December.

The climada model (Bresch and Mueller, 2017) is an open-source is a natural catastrophe (NatCat) damage model that is based on four modules:

1. Assets → geographical distribution of houses/people etc. This is created from a satellite nightlight image on a 10km scale for every country individually directly in climada

2. Damage functions → The default damage function from the *winterstorm_europe* module is heuristically adapted to the wind speed values of System 4 and ERA-Interim so that the losses are at least within a reasonable order of monetary magnitude for intense winterstorm events($\sim 10^9$ USD). As the calibration of

the damage function is neither our expertise nor possible due to the lack of actual loss data we scale all losses to the ERA-Interim through dividing by the maximum loss year observed in ERA–Interim. Thus the loss figures are ratios relative to the maximum loss in the reanalysis. For the sake of simplicity we also use the same damage function for all four countries.

3. Hazards → climada is used to transform the windstorm footprints tracked with WiTRACK (see above) into hazard sets that can subsequently be used by climada for loss calculations (via an adaptation of the *climada_ cosmo2hazard function*)

4. Adaptation measures → Not used for this study

After the iteration of steps 1.–3. we obtain an absolute annual expected loss (scaled to the maximum annual loss of ERA-Interim for the respective country) for all four countries for every year and all 51 ensemble members. Thus the loss will be presented as fractions of the costliest year in ERA-Interim.

Climada as a tool offers a lot more functions, however as the scope of our study is simply to investigate the feasibility of creating a hazard set from ensemble predictions, we limit the usage of climada to simple annual loss calculations. For more details on all the capabilities of climada the reader is referred to the climada manual (Bresch and Mueller, 2017).

The return level plots (Chapter 6.3.1) were created fitting a Generalised Pareto distribution (GPD) to the seasonal forecast ensemble with the help of the R package ismev (Gilleland and Katz, 2016). In order to investigate the proposed relationship

between the intensity/frequency of European windstorms and large-scale indices we conduct a composite analysis and check whether the phase of the NAO (or other indices) has a significant impact on the windstorm-associated losses. A positive phase of a respective index is defined as exceeding the 95th percentile of all years across all 51 ensemble members. Likewise a negative phase of an index is defined as being below the 5th percentile for all years and ensemble members. Thus 82 years out of the entire data set are classified as being positive/negative respectively.

6.3 Results

6.3.1 Estimated losses from System 4

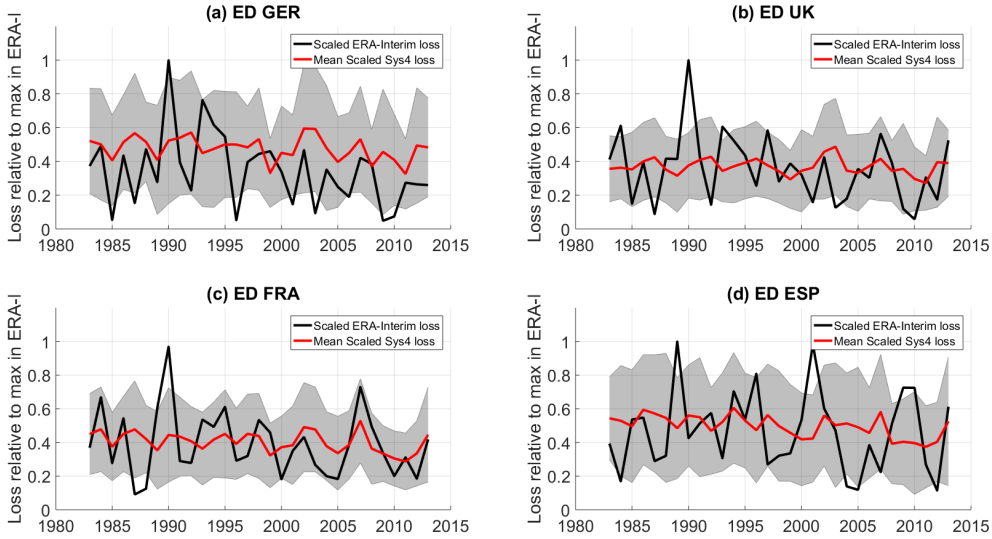


Figure 6.3.1: *Expected damage (ED) calculated with climada for (a) Germany, (b) UK, (c) France and (d) Spain in ERA-Interim (black) and System 4 (red). The standard deviation of the ensemble of System 4 is given in grey shading. All values are scaled with the maximum of ERA-Interim.*

Figure 1 shows annual losses for the four countries both for System 4 and ERA-

Interim windstorms. The grey shading represents the standard deviation for the 51 ensemble members whereas the red line represents the mean loss over all 51 members. All values are scaled to the maximum annual loss in every region. The year 1990 represents the most loss-intensive year for the UK, Germany and France. As evidenced by MunichRe (2007) this is related to the series of windstorms that hit Europe in January and February 1990 (e.g. windstorm Daria; Heming, 1990).

Years featuring other prominent windstorms like Lothar (Rivière et al., 2010) in 1999, Jeanette (Parton et al., 2009) in 2002 or Kyrill (Fink et al., 2009) in 2007 also show an above average loss. The highest relative loss in Spain was estimated for the years 1989 and 2001. Although no major storm hit Spain in 1989 the season was one of the stormiest in the recent passed for Spain including e.g. an average wave height of 7.8m was recorded in the Southwest of Spain (Rangel-Buitrago and Anfuso, 2013). The years 2009 and 2010 also mark years with extensive losses for Spain. These losses were most likely caused by windstorms Klaus and Xynthia (Lumbroso and Vinet, 2011). Xynthia is particularly interesting as it occurred during an extreme negative phase of the NAO (see section below). For all other three countries the year 2010 was amongst the least intensive loss years. Just as a rough guide, losses for 1990 in ERA-Interim as calculated with our "arbitrary" damage function come to 15 Billion USD for Germany and around 7.5 Billion USD for the UK. As mentioned before these numbers are not to be seen as real world values, thus we apply the scaling to the maximum.

The mean of the annual loss as calculated by System 4 shows a reduced variability compared to ERA-Interim. This is in line with the findings of Walz et al.

(2018b) who showed that the seasonal extreme wind speeds of System 4 feature a reduced variability compared to the observations. The inter-annual variability of ERA-Interim is nicely captured, however, within the standard deviation of the system 4 ensemble. This means that System 4 correctly spans the "loss space" of reality. The mean loss over the entire period agrees well between ERA-Interim and System 4 for the UK (0.36 vs 0.37), France (0.42 vs 0.40) and Spain (0.47 vs. 0.50). The mean loss calculated for Germany however differs considerably (0.35 vs. 0.48). Germany is also the country where the spread of the ensemble is the largest. The smallest is apparent for the UK and France. This is in line with the extreme values of the ensemble distribution: The maximum annual loss generated by System 4 for Germany is more than double the loss estimated for 1990 (2.14) whereas the maximum for the UK is around 1.34 times the 1990 loss (1.31 times 1988 loss for France and 1.86 times 1989 loss for Spain). The inter-annual variability of ERA-Interim losses for Germany is well in line with Leckebusch et al. (2007). Although they were using the cubic exceedance of the 98th percentile of local wind speeds as a proxy for loss and using ERA40 the main loss years are the same than in our study (1984, 1990, 1998).

The panels in Figure 2 depict the return level plots for Germany and the UK created via a GPD. From the plot it becomes evident that losses in Germany are considerably higher compared to the UK. The loss of 1990 (value of 1.0) for example is expected to happen within a return period of around 20 years whereas for the UK the same magnitude of loss represents a hundred year event. This is roughly the same return period for which a loss of 1.5 times the 1990 losses would be expected for Germany.

Della-Marta and Pinto (2009) estimated the return period of Daria between 24 and 39 years. Although an entire loss season is not easy to relate to one storm alone, their estimate fits well for the loss return period for Germany. The higher return period for the UK can be potentially be explained by the additional loss-intensive storms in 1990. The dashed grey lines in Figure 6.3.2 depict the uncertainties of the return levels if only calculated from ERA-Interim. Evidently the uncertainty of potential losses can be estimated considerably more accurately via System 4. The return levels of System 4 for Germany are almost completely outside of the range of ERA-Interim which means that when using ERA-Interim only, the potential loss would be severely underestimated. Thus, according to our results return periods of losses calculated from ERA-Interim should be treated with care.

The available data of only around 40 years is simply not enough to precisely benchmark for example a 200–year loss event. The uncertainty in the tail becomes too large. The loss estimates of climada naturally are very sensitive to the input data. Thus, there is no absolute truth here as the return periods calculated from System 4 could also feature a bias towards too high wind speeds. However at least the uncertainty in the tail can be reduced so that a sharper "loss space" can be obtained. The panels in Figure 6.3.2 depict the return level plots for Germany and the UK created via a GPD. From the plot it becomes evident that losses in Germany are considerably higher compared to the UK. The loss of 1990 (value of 1.0) for example is expected to happen within a return period of around 20 years whereas for the UK the same magnitude of loss represents a hundred year event. This is roughly the same return period for which a loss of 1.5 times the 1990 losses would be expected

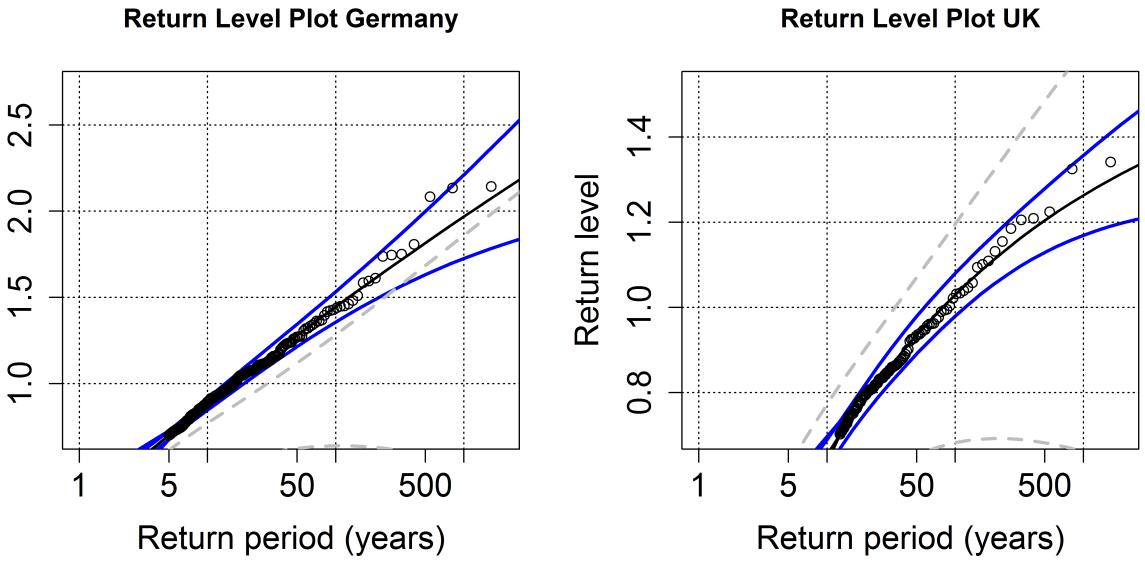


Figure 6.3.2: *Return periods for Germany (left) and the UK (right) calculated from the entire ensemble of System 4. Note the different y-axis scales.*

for Germany

6.3.2 Estimated losses linked with large-scale driver indices

The relationship between three large scale indices and the annual estimated loss for System 4 is investigated via a composite analysis. Table 1 presents the results thereof. To check for significance of deviation from choosing 82 random years, we performed a bootstrap sampling using k=100,000 samples. This bootstrapping is performed to show how loss numbers in years with extreme indices deviate from randomly sampled years. See an example for the results of the bootstrap for the NAO case in Figure 3.

As previously shown by various studies (e.g. Donat et al., 2010) the NAO has a significant impact on windstorm induced losses across Europe. The losses in Germany

during the 82 strongly positive NAO phase years are significantly higher than the mean across all 1632 (virtual) model years. The reduced loss during the negative phase of the NAO is even more significant. The same result is apparent for the UK where there is also significant more (less) windstorm related loss during a positive (negative) NAO phase. The signal for France by contrast is not as strong. There is reduced loss during the NAO positive NAO phase, however the loss during the negative phase is also lower than for the mean across all years and ensembles. Studies have shown that the NAO in seasonal forecasts can be predicted with significant skill (e.g. Scaife et al., 2014 or O'Reilly et al., 2017). As a result this would mean that a seasonal forecast exhibiting extreme NAO values for a season bears the potential of either above or below average windstorm losses. Thus an information gain regarding loss potentials. There are also reduced losses for France during a positive SCA phase. Overall there is little signal for the SCA pattern in System 4.

The loss during the positive NAO phase for Spain is significantly lower compared to the entire mean. This is in line Walz et al. (2018a) who show that there is a negative link between the NAO phase and the storminess for the Iberian Peninsula. There is not much of a significant link for the rest of the large-scale drivers bar the EA index that features some significance for the losses in Spain. This is again in line with the findings of Walz et al. (2018a) who could show that the EA pattern is a significant driver for windstorm clustering and they could confine the area of impact of the EA pattern to the East Atlantic and the northern Spain (cf. their Figure 3a). Overall there seems to be a strong link between the NAO and winter windstorm losses within System 4. The link between losses and the other two indices does appear not to be significant.

Table 6.3.1: *Results of the composite analysis presented per country and positive (>95th percentile)/negative (<5th percentile) phase of the respective index. A + corresponds to a 90%, a * corresponds to 95% and ** to 99% significance of deviation from picking 82 random years. Again the losses are relative to the most extreme loss year in ERA-Interim.*

Country	*NAO+*	*NAO-*	*EA+*	*EA-*	*SCA+*	*SCA-*	*Mean*
Germany	**0.54***	**0.37****	0.44	0.49	0.47	0.48	0.48
UK	**0.43****	**0.27****	0.38	0.37	0.36	0.39	0.36
France	**0.31***	0.35	0.40	0.38	**0.36**$^{+}$	0.40	0.40
Spain	**0.26****	0.51	**0.56***	**0.44**$^{+}$	0.47	0.50	0.49

As a result this would mean that a seasonal forecast exhibiting extreme NAO values for a season bears the potential of either above or below average windstorm losses. Thus an information gain regarding loss potentials. There are also reduced losses for France during a positive SCA phase. Overall there is little signal for the SCA pattern in System 4. This is somewhat curious as the SCA pattern has been shown to have a significant impact on windstorms (Walz et al., 2018a or Mailier et al., 2006). Walz et al. (2018b), however could show that especially the SCA pattern within System 4 looks considerably different to reality. The loss during the positive NAO phase for Spain is significantly lower compared to the entire mean. This is in line Walz et al. (2018a) who show that there is a negative link between the NAO phase and the storminess for the Iberian Peninsula. There is not much of a significant link for the rest of the large-scale drivers bar the EA index that features some significance for the losses in Spain. This is again in line with the findings of Walz et al. (2018a) who could show that the EA pattern is a significant driver for windstorm clustering and they could confine the area of impact of the EA pattern to the East Atlantic and the northern Spain (cf. their Figure 3a). Overall there seems to be a strong link between the NAO and winter windstorm losses within System 4. The link between

losses and the other two indices does appear not to be significant.

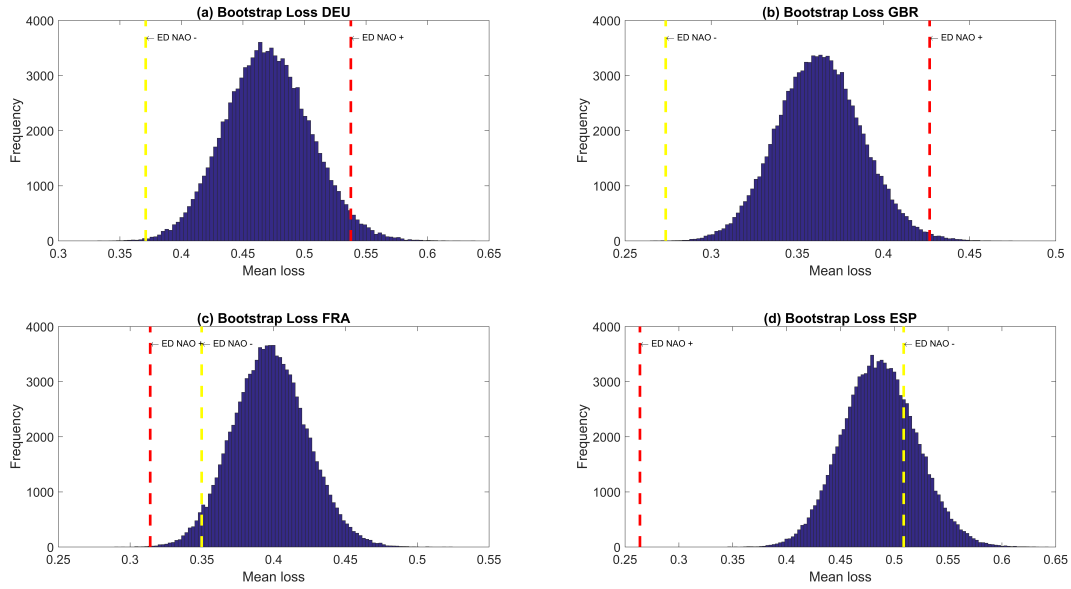


Figure 6.3.3: *Example of the bootstrapping expected damage (ED) distributions for (a) Germany, (b) UK, (c) France and (d) Spain. The mean loss for the positive (negative) NAO phase is noted with a red (yellow) vertical line*

6.4 Summary and Discussion

This study proposes the utilisation of the ECMWF System 4 hindcast data as a hazard event set for winter windstorms in Europe. The windstorm events were identified for every year (December-March) and ensemble member using the wind-based tracking tool WiTRACK. The loss calculation was realised with the open-source NatCat damage model climada directly for the tracked wind fields. In order to compare the estimated losses for System 4 they were related to losses calculated from ERA-Interim over the same period. The overall losses agree well in their magnitude. As expected, the inter–annual variability of the ensemble mean is visibly smaller than

in the observations. This is in line with Walz et al. (2018b) who show a similar results for a seasonal extreme wind speed metric. The standard deviation of the loss ensemble does capture the inter-annual variability nicely, however.

In terms of observed loss the year 1990 (e.g. storm Daria) was the most loss-intensive year for both the UK, Germany and France. The maximum for Spain in 1989 is a bit curious, however Rangel-Buitrago and Anfuso (2013) find 1989 to be one of the stormiest years with regards to wave height. The potential extreme losses in the System 4 event set differ considerably for the considered countries: The largest loss year in System 4 for Germany is more than double the 1990 losses whereas the largest loss year for the UK is "only" 1.34 times the loss in the year 1990. This is confirmed by the return level plot for Germany that shows considerably higher return levels compared to the UK equivalent. This means that years with double the loss amount of 1990 are physically possible; given that the model is a faithful representation of the atmosphere of course. The return level plot also nicely shows the more accurate estimation of loss uncertainties when utilising System 4 as a hazard set compared to uncertainties calculated from observations.

In accordance with various studies (Donat et al., 2010) the NAO is found to have a significant impact on the annual winter storm losses in Europe, especially for Germany, the UK and Spain. Our results are well in agreement with the literature showing increased (decreased) loss potentials for strongly positive NAO phases in the UK and Germany (Spain). The result for Spain is particularly striking and in line with Walz et al. (2018a) who showed a negative correlation between the NAO and windstorm occurrence for the Iberian Peninsula. This can be explained by the

position and strength (weakness) of the baroclinic jet stream during positive (negative) NAO phases and thus little (enhanced) storm activity in southern Europe. Muñoz-Díaz and Rodrigo (2003) for example found increased precipitation in Southern Spain during negative NAO years. Except for the EA pattern in Spain the other indices did not appear to have a significant impact on the potential loss. The SCA pattern which has been shown to have a significant impact on European storminess (Walz et al., 2018a or Mailier et al., 2006) does not seem to be linked with losses in System 4 except for a small signal in France. This could be in line with Walz et al. (2018b) who found the SCA pattern in System 4 to be fairly different compared to the reanalysis.

In this paper we have demonstrated that the ECMWF System 4 provides a physically consistent and realistic hazard event set which can be used for loss estimation and a more accurate estimation of loss return levels as shown by the uncertainties in Figure 2. The question posed in the title could be answered by return period estimates of 1990 losses between 20-25 years for Germany and 50-100 years for the UK. We could identify a strong link between the NAO and losses for Germany and the UK in particular. This could prove to be vital information regarding future runs of seasonal forecasts as the there is a significantly larger chance of more loss occurring if the NAO is extremely positive in System 4, thus an information gain regarding future loss potentials.

Chapter 7

Synthesis

"After a storm comes a calm"

– Matthew Henry, English Clergyman

7.1 Summary

This thesis is dedicated to advance the understanding of European winter windstorms in five different aspects:

- Quantifying the severity of windstorms in a climatological sense → *Quantification*
- Identifying large–scale drivers linked with serial clustering of windstorms → *Temporal variability*
- Assessment of the predictability of extreme winds associated with windstorms and identification of drivers of extreme winds → *Predictability*
- Characterisation of windstorm tracks and identification of most extreme windstorm clusters → *Spatial variability & maximum intensity*
- Creating a physically consistent windstorm hazard set to assess the uncertainty in losses for Europe using a NatCat model → *Impact*

Regarding the quantification of windstorms chapter 2 presents a way to objectively quantify the extremity of windstorms in a climatological sense. The work of this chapter was published in Walz et al. (2017). Compared to the SSI (Leckebusch et al., 2008a) which is designed to estimate the severity of a storm in an impact related manner, the DI-SSI provides an estimation that compares climatological extremeness. As the SSI does not account for the tail of the distribution, large SSI values are observed in regions that are less exposed to extreme wind speeds (see Table 2.3.2). As the resilience towards extreme winds in those areas is potentially lower the larger SSI values are justified (see also Figure 5.4.3 the relative low SSI values for the British Isles). From a climatological perspective, however, it also

makes sense to have an indication on the severity of a storm regardless of its geographical location. Thus having a single number that quantifies the extremity of the storm with regard to the tail of the local wind speed climatology. This was successfully accomplished by the development of the DI-SSI. By making use of a GPD distribution and a equiprobability transformation, any exceedance of the 98^{th} is assigned the same value. This changes the severity ranking of prominent European windstorm episodes considerably (Table 2.3.2). Storms like Klaus or Martin that affected mostly Central or Souther Europe end up with a lower severity ranking whereas storms like Daria, Anatol or Kyrill climbed in the relative severity ranking. The case study of Storm Klaus (Figure 2.3.2) unveils the difference between the two approaches: Whereas the largest SSI contributions occur solely over land, the DI-SSI classifies the winds over the ocean close the coast as the most extreme ones. The correlation analysis between the SSI/DI-SSI and the NAO showed a more homogeneous correlation pattern for the DI-SSI especially in regions outside of the classic North Atlantic stormtrack. The correlation between the windstorm frequency and the two severity time series draws a similar picture – a larger and more homogeneous correlation pattern across the North Atlantic domain. These two results confirm that the DI-SSI developed in chapter 2 has the potential to be used as a standard metric to quantify the meteorological extremeness of European windstorms regardless of their geographical occurrence.

The temporal variability of windstorms was investigated in chapter 3. This chapter consists of two parts:

- Impact perspective – given a set of regions, what are the significant large–scale drivers to explain the most inter–annual variability?

- Physical perspective – given a set of large–scale drivers, where is their main area of influence in the North Atlantic/European domain?

The results of this chapter were published in Walz et al. (2018a). The first question was addressed by the development of a Poisson model to statistically model the serial clustering of windstorms (chapter 3.3). Among the identified significant drivers were prominent the NAO or the temperature difference between North America and the West Atlantic that had previously been linked with European storminess (Donat et al., 2010, Mailier et al., 2006, Wild et al., 2015). A striking feature of this analysis was the importance of the SCA pattern for European storminess however for it appeared as the most influential large–scale driver for five out of seven regions (Table 3.4.1). The SCA pattern is associated with a blocking situation across Central Europe that prevents windstorms from affecting countries like France or Europe. This is an interesting result as evidently it is not sufficient just to examine individual large–scale patterns to deduce the potential storminess of a winter season. The Poisson model for modelling the serial clustering showed considerable skill explaining up 70% of the inter–annual variance for the British Isles (Table 3.4.2). The average explained variance for all seven regions is larger than 50%, thus just by using a set of atmospheric/oceanic drivers more than half of the variability of European winter windstorms can be explained. This can be particularly interesting for the actuarial sector. The map of drivers (Figure 3.4.2) provides a descriptive overview over the areas of impact of the five most influential drivers identified by the Poisson model. In line with the results from the model, the SCA pattern is the driver that explains most of the variability for Central Europe whereas the impact of the NAO is restricted to the British Isles and the North Atlantic. It has to be kept in mind, though, that only the "winning" driver is plotted on the map

of drivers, thus the NAO might be the second most important driver for areas in Central Europe. As also identified by Vitolo et al. (2009) or Mailier et al. (2006), the EA pattern is of significant importance in the East Atlantic, Western France and Northern Spain. Overall this chapter provides a comprehensive overview on the importance of large–scale drivers for the inter–annual variability of European windstorms and serial clustering in particular.

Large–scale drivers have an important effect on European winter storminess as seen from chapter 3. Chapter 4 tried to link the large–scale dynamics of the European/North Atlantic domain with the predictability of extreme wind speeds associated with winter windstorms. This chapter was published in Walz et al. (2018b). The study was trying to answer four main questions:

- How well are large–scale patterns represented in seasonal forecast products (ECMWF System 4 in particular) compared to reanalysis?
- Similar to chapter 3: What are the large–scale drivers explaining the most variability of extreme winds for a winter season?
- What is the skill/predictability of extreme wind speeds associated with windstorms in System 4?
- Can large–scale conditions in autumn help to predict the potential storminess of an upcoming DJF season?

As it turns out the large–scale patterns in System 4 look substantially different to the ones in the observations. The prominent NAO dipole pattern as a result from an EOF analysis appears shifted downstream in System 4 so that the main centres of action are located over Scandinavia and Southern Europe. This is in contrast

to the reanalysis in which the dipole is in its expected location over Iceland and the Azores. The downstream shift of the MSLP variability patterns in System 4 appears to be of systematic nature as also for the PNA pattern the centres of action are moved downstream. (Figure 4.4.2). Other patterns such as the EA pattern look substantially different in System 4. The second step in the analysis of System 4 was to examine what drives the extreme wind speeds on an inter–annual time scale. Due to the large discrepancies between the EOF patterns of System 4 and ERA–Interim the EOF eigenvectors of ERA–Interim were used to calculate the variability time series (NAO, PNA, ...). The resulting maps (Figure 4.4.4) are analogous to the "map of drivers" presented in Figure 3.4.2 in chapter 3. Although the System 4 MSLP anomalies were projected onto the ERA–Interim based eigenvectors, the resulting maps looked substantially different. The drivers of extreme wind speeds for ERA–Interim draw a similar picture to the inter–annual variability of windstorm counts in chapter 3. The NAO, and the EA pattern appear as the main drivers of extreme wind speeds for the European core winter season (DJF). The same map produced with the System 4 MSLP anomalies showed a clear difference to the reanalysis–based one: Instead of the NAO, the SCA pattern seems to be the most influential driver for Central Europe. This would be in accordance to the result in chapter 3 it has to be kept in mind, however, that the variable in question was amount of windstorms whereas for this analysis the magnitude of wind speeds over an entire season was investigated. Other drivers especially the ones expected for the Pacific region also appear as drivers in the European domain which is most likely an artefact of System 4. The large–scale drivers related to extreme wind speeds generally seem to be poorly represented in System 4 compared to the reanalysis. This becomes even more obvious when looking at the drivers produced by the System 4 based

EOF eigenvectors. The patterns seem to have shifted completely so that the PNA appears as the main driver of Euorpean storminess in System 4. As this analysis is not necessarily a fair comparison, these figures can only be seen as a trial and are presented in the appendix.

The predictive skill of System 4 was assessed with a more unconventional skill score that has its roots in information theory. Generally System 4 added little additional information to the climatology within the main North Atlantic stormtrack, however there are pockets of predictive skill for parts of Scandinavia, Eastern Canada and the Central Atlantic for all three analysed percentiles. Thus, System 4 provides more information than the climatology for those regions. The last question above was addressed by developing a multi–linear regression model using autumn conditions to predict the upcoming DJF season. The results were surprisingly positive, especially for the Central European region where the statistical model showed more skill than System 4. This could be useful information for companies associated with wind storm risk modelling, e.g. the actuarial sector.

The spatial variability and maximum intensity of windstorms their associated cyclones was analysed using the probabilistic clustering approach proposed by Gaffney et al. (2007). In order to have a large sample size of potential windstorms, System 4 was used as a pool of potential "alternative realities". Thus the 51 members each comprising 32 years were used as a climate archive resulting in more than 1500 years of data. Based on Gaffney et al. (2007) and empirical analysis the windstorms were classified in three different clusters (Figure 5.4.1 and 5.4.2). In total more than 11,000 storms could be identified for the British Isles region and more than 14,000 for the Central European region. This large sample size made it possible to make

sound estimates on the return period of the SSI values per region and cluster. It turns out that windstorms categorised in Cluster 1 harbour the largest hazard potential in terms of SSI values. This is most likely due to the storms in Cluster 1 being the largest and longest ones on average. Cluster 3 contains storms that approach Europe from a north-westerly direction. Storms featured in Cluster 3 tend to bear the lowest immediate hazard potential, however due to their track they harbour the risk of storm surges, especially for the German Bight (e.g. Befort et al., 2015). The meteorological conditions present during the three different clusters of windstorms revealed an enhanced dipole structure between the Icelandic Low and the Azores high. Depending on the cluster the anomaly structure was shifted more downstream. In particular for Cluster 3 for which the tracks follow geostrophic conditions on the backside of the wave. A more dynamical explanation for the different tracks could also be given by the PV gradient (e.g. Hoskins et al. (1985)) in the motion of a windstorm. Rivière et al. (2012) could show that the PV was responsible for the poleward movement of windstorm Xynthia perpendicular to the mean flow conditions. Future research could entail a detailed analysis of the PV gradient conditions for windstorms in each respective cluster (see Chapter 7.2).

The extreme value analysis for the ETCs confirmed the findings on the windstorms: Cluster 1 contains cyclones with the lowest core pressure and the largest curvature values across all return periods. Both of these variables are a good proxy of strong winds associated with the cyclone (e.g. Murray and Simmonds, 1991). The minimum potential core pressure/maximum curvature, however is more extreme for Cluster 2 and 3 (e.g. British Isles: 911 hPa/6.2 hPa/deg.lat^2 for Cluster 1 vs. 909 hPa/6.4 hPa/deg.lat^2 for Cluster 2). These numbers are well below/above the 1000–

year return period value for both of the variables. Hence they can only be seen as a "physical extreme" rather than a realistic value to be observed associated with an actual ETC. Overall this chapter proposed a way to utilise seasonal ensemble forecast systems as climate archives. As an application of this archive the spatial variability of windstorms and ETCs was investigated. The classification in three different clusters revealed considerable differences in windstorms/ETCs especially with regard to the extremes and potential "worst–case" storms.

The last chapter of this thesis dealt with the impact of winter windstorms. Similar to chapter 5 System 4 was used as a pool of alternative realities as observations of extreme windstorms is limited roughly to the last hundred years. Due to the set–up of 51 ensemble members System 4 is comprised of more than 1500 years of model data. Wind fields associated with windstorms were identified in order to generate a hazard event set of potential loss–intensive storms. Likewise storms were identified in ERA–Interim in order to compare the climatologies of annual losses. The loss calculation was done with the open source NatCat loss model *climada* (Bresch and Mueller, 2017). All losses were scaled to the maximum loss in the observations in order to get a feeling what "loss space" System 4 can generate. The results show that losses of a double magnitude compared to the observed losses in 1990 (Daria, Vivian...) are physically possible for Germany. The maximum loss in System 4 for the UK for example is "only" 1.34 times larger than the loss in 1990. The inter–annual variability of the ensemble mean of the losses is lower compared to the reanalysis. This is in line with the findings in chapter 4 where it was proposed that the ensemble mean of extreme wind speeds in System 4 do not show a large inter–annual variability. Losses in the range of the year 1990 would be a 25-30–year event for

Germany whereas they represent a 100–year event for the UK. Results from Chapter 5 show that depending on the cluster a 100–year event for the British Isles with regards to minimum core pressure ranges between 928–935 hPa. The lowest core pressure recorded for windstorm Daria was 949 hPa (McCallum and Norris, 1990), thus around 10 hPa higher than an estimated 100–year event with regards to loss. When comparing these numbers it has to be kept in mind, however, that the 100–year event with regards to loss is associated with the entire year of 1990 and not just one particular storm. The lowest pressure for windstorm Vivian for example was estimated to be 940 hPa (McCallum and Norris, 1990), thus closer to a 100–year event for the UK and around a 50–year event for Germany, where the windstorm was more devastating than in the UK. The 50–year event for Germany is in a similar order to the 25–30–year loss event for Germany estimated in Chapter 6.
The variability of the observations is captured nicely by the spread of the entire ensemble, though. The link between the NAO and windstorm losses was shown to be significant in System 4 for Germany, the UK and Spain. The SCA and the EA pattern, however did not show much association with losses bar the EA pattern for Spain. These findings are well in agreement with literature (Donat et al., 2010, Mailier et al., 2006, Pinto et al., 2009) and also chapter 2 and 3 (Walz et al., 2017, 2018a).

The presented thesis provides an extensive study of European winter windstorms: It investigates the physical drivers that lead to temporal variability (chapter 3), quantifies windstorms both in a climatological and loss–related manner (chapters 2 and 6), estimates maximum potential intensities (chapter 5) and last but not least examines the predictability of extreme wind speeds associated with windstorms (chapter 4).

Essentially it tries to answer all the questions that Captain Richards and Urbain Le Verrier would have asked back in 1854 and surely they would have been happy to have access to this kind of information back in the day. Thus in a way the motivation for this thesis has been building up over the last 160 years...

7.2 Outlook

The current thesis with its five chapters represents a comprehensive study to advance the understanding of quantification, variability, predictability and extremeness of European winter windstorms. Despite the substantial advances presented here, there is naturally a potential for future work regarding the four topics discussed.

So far the DI-SSI has only been applied to European windstorms, however it would be intriguing to see how it would perform for tropical cyclones (TCs). Due to the parametric nature of the index it could potentially help to build a "severity climatology" for hurricanes or typhoons. As the SSI is a more impact related quantification it would be interesting to see how major TCs compare in their severity values similar to what was done for windstorms in Table 2.3.2. A climatological severity could also be of interest to build a link function between the Saffir–Simpson scale (Simpson and Saffir, 1974) and the DI-SSI. Another idea would be to use the DI-SSI not only to quantify wind speeds, but also for precipitation associated with windstorms. A combination of both (wind and precipitation) would also be an option in order to quantify the compound hazard of extreme wind and rainfall. Technically the DI–SSI could also be used for other variables unrelated to windstorms, however one would need to find a way to "track" a certain event, e.g. a heatwave. Similar ideas were already proposed by Kruschke (2015) who argued that the WiTRACK algorithm currently only used for windstorm tracking could also be used to track other extreme meteorological variables (i.e. geopotential).

The statistical model developed for modelling the serial clustering of windstorms in its current version can only be used in "real time" as its training data set (hence all

the large–scale drivers) as well as the windstorm counts to be predicted are during DJF. Thus, the only way to use the model in "forecast–mode" would be to use it with seasonal forecast data which is issued for example in November (e.g. System 4). If the training data set, however, comprised of SON one could build a model to use the observed large–scale drivers of autumn to make a prediction about the upcoming DJF season (lagged model). The skill scores would most likely be lower than in Table 5.4.2, the model could still provide useful information on the windstorm potential, however. Even if the lagged model was not even used for a prediction, one could still learn about drivers in autumn that significantly impact the nature of the upcoming DJF season. So far any interactions between the predictors were disallowed. This would also bear a potential for improvement as some of the drivers might not necessarily have an impact on their own, if combined with another driver, however there might be a significant contribution to the variability. This could be the case for interactions between SST related drivers and atmospheric ones in particular.

As apparent from Figure 4.4.1 and 4.4.2 there seems to be a downstream shift of the MSLP variability patterns. In order to see if there is a lead time dependent shift it would be of interest to repeat the analysis for each individual lead month and to see if there is a temporal progression of this downstream shift with increasing lead time. The analysis on seasonal predictability of extreme wind speeds so far only comprises 32 years of data. Weisheimer et al. (2017) introduced century–long hindcasts which are based on System 4 which could be used for a more robust analysis if combined with a century–long reanalysis (e.g. ERA–20C). That way some of the noise in determining the large–scale drivers could potentially be reduced. Additionally some of the decadal variability of the large–scale drivers might be picked up on for the

analysis. More generally the identification of large–scale drivers could be extended to other variables (e.g. temperature or geopotential height anomalies).
The spatial clustering in chapter 5 could be conditioned on different NAO phases. This would make it possible to estimate an occurrence probability of a cluster depending on the prevailing NAO phase. The analysis could also be repeated for TCs. This was already done by Camargo et al. (2007), however their data set is only comprised of around 50 years of observational data. By using System 4 as a climate archive analogous to what was done in this thesis a huge hazard set of TCs could be produced. The WiTRACK algorithm which would be used for identifying the TCs would also provide SSI values and other attributes of each individual storm. So similar assessments in terms of return periods would be possible. Instead of using the SSI value as a quantification of the severity it would also be possible to use the DI-SSI developed in chapter 2. This would most likely affect the results in Figure 5.4.4. As discussed in chapter 2 Central Europe generally features larger SSI values due to the definition of the SSI. By using the DI-SSI one could make a more climatological assessment of the severity of the windstorms in each cluster for the four different regions. A more dynamical analysis of the windstorms in each cluster might answer some questions why certain tracks (or more generally the clusters) behave the way that they do. Windstorms in each respective cluster could be analysed for latent heat release using a PV analysis similar to Tierney et al. (2018). That way PV anomalies could be attributed to each cluster. The PV anomalies could then further be used to associate windstorms with a specific cluster.

As the last chapter on the impact of windstorms is lined out as a feasibility study it contains the potential for further analysis. The damage function that was used to

estimate the annual losses for example was kept the same for all four countries. For an operational loss model the vulnerability of every country or even for every assets would need to be determined individually. Additionally it would be worthwhile to use wind gusts rather than the 10m wind speeds as the hazard variable. The wind fields obtained from WiTRACK would then be used to cut out the respective gust wind field for a storm. For System 4 however the gusts are only available every 24 hours so that for short events there would only be one available time step for an event compared to four time steps for the 10m wind. As it was shown that System 4 generates a realistic "hazard space" the approach could be transferred to other atmospheric perils like TCs. *Climada* also includes a TC module so that the loss calculation for the TC event set could also be performed fairly easily.

Bibliography

Agresti, A. (2003). *Categorical data analysis*, volume 482. John Wiley & Sons.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Ambaum, M. H., Hoskins, B. J., and Stephenson, D. B. (2001). Arctic oscillation or North Atlantic oscillation? *Journal of Climate*, 14(16):3495–3507.

Anderson, D., Stockdale, T., Balmaseda, M., Ferranti, L., Vitart, F., Molteni, F., Doblas-Reyes, F., Mogensen, K., and Vidard, A. (2007). Development of the ECMWF seasonal forecast System 3. *ECMWF Technical Memoranda*, 503.

Athanasiadis, P. J., Bellucci, A., Scaife, A. A., Hermanson, L., Materia, S., Sanna, A., Borrelli, A., MacLachlan, C., and Gualdi, S. (2017). A multisystem view of wintertime NAO seasonal predictions. *Journal of Climate*, 30(4):1461–1475.

Baldwin, M. P. and Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, 294(5542):581–584.

Bärring, L. and von Storch, H. (2004). Scandinavian storminess since about 1800. *Geophysical Research Letters*, 31(20).

Beaufort, F. (1805). Beaufort wind scale. *British Rea-Admiral*.

Befort, D., Fischer, M., Leckebusch, G., Ulbrich, U., Ganske, A., Rosenhagen, G., and Heinrich, H. (2015). Identification of storm surge events over the German bight from atmospheric reanalysis and climate model data. *Natural Hazards and Earth System Sciences*, 15(6):1437–1447.

Befort, D., Wild, S., Knight, J., Lockwood, J., Thornton, H., Hermanson, L., Bett, P., Weisheimer, A., and Leckebusch, G. (2018). Seasonal Forecast Skill for Extra-tropical Cyclones and Windstorms. *Quarterly Journal of the Royal Meteorological Society*.

Befort, D. J., Wild, S., Kruschke, T., Ulbrich, U., and Leckebusch, G. C. (2016). Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses. *Atmospheric Science Letters*, 17(11):586–595.

Blender, R., Fraedrich, K., and Lunkeit, F. (1997). Identification of cyclone–track regimes in the North Atlantic. *Quarterly Journal of the Royal Meteorological Society*, 123(539):727–741.

Bloomfield, H., Shaffrey, L., Hodges, K., and Vidale, P. L. (2018). A critical assessment of the long-term changes in the wintertime surface Arctic Oscillation and Northern hemisphere storminess in the ERAGermanC reanalysis. *Environmental Research Letters*, 13(9):094004.

Branstator, G. (2002). Circumglobal teleconnections, the jet stream waveguide, and the North Atlantic Oscillation. *Journal of Climate*, 15(14):1893–1910.

Bresch, D. and Mueller, L. (2017). Climada manual.

Brönnimann, S., Annis, J. L., Vogler, C., and Jones, P. D. (2007). Reconstructing the

quasi-biennial oscillation back to the early 1900s. *Geophysical Research Letters*, 34(22).

Bueh, C. and Nakamura, H. (2007). Scandinavian pattern and its climatic impact. *Quarterly Journal of the Royal Meteorological Society*, 133(629):2117–2131.

Camargo, S. J., Robertson, A. W., Gaffney, S. J., Smyth, P., and Ghil, M. (2007). Cluster analysis of typhoon tracks. part i: General properties. *Journal of Climate*, 20(14):3635–3653.

Charney, J. G. (1947). The dynamics of long waves in a baroclinic westerly current. *Journal of Meteorology*, 4(5):136–162.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer, London.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al. (2011). The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137(654):1–28.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Cusack, S. (2013). A 101 year record of windstorms in the Netherlands. *Climatic change*, 116(3-4):693–704.

Czaja, A. and Frankignoul, C. (1999). Influence of the North Atlantic SST on the atmospheric circulation. *Geophysical Research Letters*, 26(19):2969–2972.

Czaja, A. and Frankignoul, C. (2002). Observed impact of Atlantic SST anomalies on the North Atlantic Oscillation. *Journal of Climate*, 15(6):606–623.

Dawson, A., Palmer, T., and Corti, S. (2012). Simulating regime structures in weather and climate prediction models. *Geophysical Research Letters*, 39(21).

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., and Bauer, P. (2011). The ERA–Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656):553–597.

Della-Marta, P. M., Liniger, M. A., Appenzeller, C., Bresch, D. N., Köllner-Heck, P., and Muccione, V. (2010). Improved estimates of the European winter windstorm climate and the risk of reinsurance loss using climate model data. *Journal of Applied Meteorology and Climatology*, 49(10):2092–2120.

Della-Marta, P. M. and Pinto, J. G. (2009). Statistical uncertainty of changes in winter storms over the North Atlantic and Europe in an ensemble of transient climate simulations. *Geophysical Research Letters*, 36(14):n/a–n/a.

DelSole, T. (2004). Predictability and information theory. Part i: Measures of predictability. *Journal of the atmospheric sciences*, 61(20):2425–2440.

DelSole, T., Nattala, J., and Tippett, M. K. (2014). Skill improvement from increased ensemble size and model diversity. *Geophysical Research Letters*, 41(20):7331–7342.

DelSole, T. and Tippett, M. K. (2007). Predictability: Recent insights from information theory. *Reviews of Geophysics*, 45(4).

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., and Rodrigues, L. R. (2013). Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4(4):245–268.

Donat, M. G., Leckebusch, G. C., Pinto, J. G., and Ulbrich, U. (2010). Examination of wind storms over central Europe with respect to circulation weather types and NAO phases. *International Journal of Climatology*, 30(9):1289–1300.

Donat, M. G., Pardowitz, T., Leckebusch, G. C., Ulbrich, U., and Burghoff, O. (2011). High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over Germany. *Nat. Hazards Earth Syst. Sci.*, 11(10):2821–2833.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical research letters*, 41(15):5620–5628.

Easterling, D. R., Evans, J., Groisman, P. Y., Karl, T. R., Kunkel, K. E., and Ambenje, P. (2000). Observed variability and trends in extreme climate events: a brief review. *Bulletin of the American Meteorological Society*, 81(3):417–426.

Economou, T., Stephenson, D. B., and Ferro, C. A. (2014). Spatio-temporal modelling of extreme storms. *The Annals of Applied Statistics*, pages 2223–2246.

Economou, T., Stephenson, D. B., Pinto, J. G., Shaffrey, L. C., and Zappa, G. (2015). Serial clustering of extratropical cyclones in a multi-model ensemble of historical and future simulations. *Quarterly Journal of the Royal Meteorological Society*, 141(693):3076–3087.

EEA (2011). Mapping the impacts of natural hazards and technological accidents in Europe.

Encyclopedia, W. (2008). Urbain Jean Joseph Le Verrier." science and its times: Understanding the social significance of scientific discovery. `https://www.encyclopedia.com/people/science-and-technology/astronomy-biographies/urbain-jean-joseph-leverrier#2830902591`. Accessed: 04/10/2018.

Fernández-González, S., Martín, M., Merino, A., Sánchez, J., and Valero, F. (2017). Uncertainty quantification and predictability of wind speed over the Iberian peninsula. *Journal of Geophysical Research: Atmospheres*, 122(7):3877–3890.

Fink, A. H., Brücher, T., Ermert, V., Krüger, A., and Pinto, J. G. (2009). The European storm kyrill in january 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change. *Natural Hazards and Earth System Science*, 9(2):405–423.

Froude, L. S. (2010). TIGGE: Comparison of the prediction of Northern hemisphere extratropical cyclones by different ensemble prediction systems. *Weather and Forecasting*, 25(3):819–836.

Froude, L. S., Bengtsson, L., and Hodges, K. I. (2007). The predictability of ex-

tratropical storm tracks and the sensitivity of their prediction to the observing system. *Monthly weather review*, 135(2):315–333.

Gaffney, S. J., Robertson, A. W., Smyth, P., Camargo, S. J., and Ghil, M. (2007). Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate dynamics*, 29(4):423–440.

Gilleland, E. and Katz, R. W. (2016). extremes2.0: An extreme value analysis package inr. *Journal of Statistical Software*, 72(8).

Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., Kharin, V., Merryfield, W., Deser, C., and Mason, S. J. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, 40(1-2):245–272.

Grassi, B., Redaelli, G., and Visconti, G. (2013). Arctic sea ice reduction and extreme climate events over the Mediterranean region. *Journal of Climate*, 26(24):10101–10110.

Greatbatch, R. J., Lu, J., and Peterson, K. A. (2004). Nonstationary impact of ENSO on Euro–Atlantic winter climate. *Geophysical research letters*, 31(2).

Hair Jr, J. F., Anderson, R. E., Tatham, R. L., and William, C. (1995). Black (1995), multivariate data analysis with readings. *New Jersy: Prentice Hall*.

Heffernan, J., Stephenson, A., and Gilleland, E. (2012). Ismev: an introduction to statistical modeling of extreme values. *R package version*, 1.

Held, H., Gerstengarbe, F.-W., Pardowitz, T., Pinto, J., Ulbrich, U., Born, K., Donat, M., Karremann, M., Leckebusch, G., Ludwig, P., Nissen, K., Österle, H.,

Prahl, B., Werner, P., Befort, D., and Burghoff, O. (2013). Projections of global warming-induced impacts on winter storm losses in the German private household sector. *Climatic Change*, 121(2):195–207.

Heming, J. (1990). The impact of surface and radiosonde observations from two Atlantic ships on a numerical weather prediction model forecast for the storm of 25 january 1990. *Meteorological Magazine*, 119(1421):249–259.

Hibbert, C. (1961). *The destruction of Lord Raglan: a tragedy of the Crimean War, 1854-55*. Longmans.

Hodges, K. I., Lee, R., and Bengtsson, L. (2011). A comparison of extratropical cyclones in recent reanalyses ERA-Interim, nasa merra, ncep cfsr, and jra-25. *Journal of Climate*, 24(18):4888–4906.

Hoskins, B. (1974). The role of potential vorticity in symmetric stability and instability. *Quarterly Journal of the Royal Meteorological Society*, 100(425):480–482.

Hoskins, B. J., McIntyre, M., and Robertson, A. W. (1985). On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, 111(470):877–946.

Hunter, A., Stephenson, D. B., Economou, T., Holland, M., and Cook, I. (2016). New perspectives on the collective risk of extratropical cyclones. *Quarterly Journal of the Royal Meteorological Society*, 142(694):243–256.

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science*, 269(5224):676–679.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical review*, 106(4):620.

Jolliffe, I. T. (1986). *Principal component analysis and factor analysis*, pages 115–128. Springer.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200.

Kim, H.-M., Webster, P. J., and Curry, J. A. (2012). Seasonal prediction skill of ECMWF System 4 and ncep cfsv2 retrospective forecast for the Northern hemisphere winter. *Climate Dynamics*, 39(12):2957–2973.

Klawa, M. and Ulbrich, U. (2003). A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Natural Hazards and Earth System Science*, 3(6):725–732.

Knight, J. R., Folland, C. K., and Scaife, A. A. (2006). Climate impacts of the Atlantic multidecadal oscillation. *Geophysical Research Letters*, 33(17).

Krstajic, D., Buturovic, L. J., Leahy, D. E., and Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics*, 6(1):1.

Kruschke, T. (2015). *Winter wind storms: Identifcation, verifcation of decadal predictions, and regionalization.* Thesis, Berlin, Freie Universität Berlin, Diss., 2015.

Kunz, M., Mohr, S., Rauthe, M., Lux, R., and Kottmeier, C. (2010). Assessment of extreme wind speeds from regional climate models – part 1: Estimation of return values and their evaluation. *Nat. Hazards Earth Syst. Sci.*, 10(4):907–922.

Kurihara, Y., Bender, M. A., Tuleya, R. E., and Ross, R. J. (1995). Improvements

in the GFDL hurricane prediction system. *Monthly Weather Review*, 123(9):2791–2801.

Lamb, H. and Frydendahl, K. (1991). *Historic Storms of the North Sea, British Isles and Northwest Europe*. Cambridge University Press.

Landsberg, H. (1954). Storm of Balaklava and the daily weather forecast. *The Scientific Monthly*, 79(6):347–352.

Leckebusch, G. C., Renggli, D., and Ulbrich, U. (2008a). Development and application of an objective storm severity measure for the Northeast Atlantic region. *Meteorologische Zeitschrift*, 17(5):575–587.

Leckebusch, G. C., Ulbrich, U., Fröhlich, L., and Pinto, J. G. (2007). Property loss potentials for European midlatitude storms in a changing climate. *Geophysical Research Letters*, 34(5).

Leckebusch, G. C., Weimer, A., Pinto, J. G., Reyers, M., and Speth, P. (2008b). Extreme wind storms over Europe in present and future climate: a cluster analysis approach. *Meteorologische Zeitschrift*, 17(1):67–82.

Liberato, M. L., Pinto, J. G., Trigo, I. F., and Trigo, R. M. (2011). Klaus–an exceptional winter storm over northern Iberia and southern France. *Weather*, 66(12):330–334.

Lim, Y.-K. (2015). The East Atlantic/West Russia (EA/WR) teleconnection in the North Atlantic: climate impact and relation to Rossby wave propagation. *Climate Dynamics*, 44(11-12):3211–3222.

Lindgrén, S. and Neumann, J. (1980). Great historical events that were significantly affected by the weather: 5, some meteorological events of the crimean war and

their consequences. *Bulletin of the American Meteorological Society*, 61(12):1570–1583.

Lloyd-Hughes, B. and Saunders, M. A. (2002). A drought climatology for Europe. *International journal of climatology*, 22(13):1571–1592.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141.

Lowe, R., García-Díez, M., Ballester, J., Creswick, J., Robine, J.-M., Herrmann, F., and Rodó, X. (2016). Evaluation of an early-warning system for heat wave-related mortality in Europe: implications for sub-seasonal to seasonal forecasting and climate services. *International journal of environmental research and public health*, 13(2):206.

Lumbroso, D. and Vinet, F. (2011). A comparison of the causes, effects and aftermaths of the coastal flooding of England in 1953 and France in 2010. *Natural Hazards and Earth System Sciences*, 11(8):2321–2333.

MacLachlan, C., Arribas, A., Peterson, K., Maidens, A., Fereday, D., Scaife, A., Gordon, M., Vellinga, M., Williams, A., and Comer, R. (2015). Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141(689):1072–1084.

MacLeod, D., Torralba, V., Davis, M., and Doblas-Reyes, F. (2018). Transforming climate model output to forecasts of wind power production: how much resolution is enough? *Meteorological Applications*, 25(1):1–10.

Madec, G. (2008). Nemo reference manual, ocean dynamic component: Nemo-opa. *Note du Pôle modélisation, Inst. Pierre Simon Laplace, Fr.*

Mailier, P. J., Stephenson, D. B., Ferro, C. A., and Hodges, K. I. (2006). Serial clustering of extratropical cyclones. *Monthly weather review*, 134(8):2224–2240.

Materia, S., Borrelli, A., Bellucci, A., Alessandri, A., Di Pietro, P., Athanasiadis, P., Navarra, A., and Gualdi, S. (2014). Impact of atmosphere and land surface initial conditions on seasonal forecasts of global surface temperature. *Journal of Climate*, 27(24):9253–9271.

McCallum, E. and Norris, W. (1990). The storms of January and February 1990. *Meteorological Magazine*, 119(1419):201–210.

McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, volume 382. John Wiley & Sons.

Mellersh, H. E. L. (1968). *FitzRoy of the Beagle*. Hart-Davis.

MetOffice (2016). National Meteorological Library and Archive Factsheet 9 — Weather extremes.

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., and Vitart, F. (2011). *The new ECMWF seasonal forecast system (System 4)*. European Centre for Medium-Range Weather Forecasts.

MunichRe (2007). Zwischen Hoch und Tief – Wetterrisiken in Mitteleuropa.

Muñoz-Díaz, D. and Rodrigo, F. (2003). Effects of the North Atlantic Oscillation on the probability for climatic categories of local monthly rainfall in southern spain. *International Journal of climatology*, 23(4):381–397.

Murray, R. J. and Simmonds, I. (1991). A numerical scheme for tracking cyclone centres from digital data. *Australian Meteorological Magazine*, 39(3).

Nissen, K., Leckebusch, G., Pinto, J. G., Renggli, D., Ulbrich, S., and Ulbrich, U. (2010). Cyclones causing wind storms in the Mediterranean: characteristics, trends and links to large-scale patterns. *Natural Hazards and Earth System Science*, 10(7):1379–1391.

Nissen, K. M., Ulbrich, U., Leckebusch, G. C., and Kuhnel, I. (2014). Decadal windstorm activity in the North Atlantic-European sector and its relationship to the meridional overturning circulation in an ensemble of simulations with a coupled climate model. *Climate Dynamics*, 43(5-6):1545–1555.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5):673–690.

Ogutu, G. E., Franssen, W. H., Supit, I., Omondi, P., and Hutjes, R. W. (2017). Skill of ECMWF System-4 ensemble seasonal climate forecasts for East africa. *International Journal of Climatology*, 37(5):2734–2756.

O'Reilly, C. H., Heatley, J., MacLeod, D., Weisheimer, A., Palmer, T. N., Schaller, N., and Woollings, T. (2017). Variability in seasonal forecast skill of Northern hemisphere winters over the twentieth century. *Geophysical Research Letters*, 44(11):5729–5738.

Osinski, R., Lorenz, P., Kruschke, T., Voigt, M., Ulbrich, U., Leckebusch, G., Faust, E., Hofherr, T., and Majewski, D. (2016). An approach to build an event set of European windstorms based on ECMWF EPS. *Natural Hazards and Earth System Sciences*, 16(1):255–268.

Palmer, T., Doblas-Reyes, F., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., Feddersen, H., Cantelaube, P., Terres, J., and Davey, M. (2004). Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, 85(6):853–872.

Palutikof, J. and Skellern, A. (1991). Storm severity over Britain. *A Report to Commercial Union General Insurance, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich (UK).*

Parton, G., Vaughan, G., Norton, E., Browning, K., and Clark, P. (2009). Wind profiler observations of a sting jet. *Quarterly Journal of the Royal Meteorological Society*, 135(640):663–680.

Philipp, A., Della-Marta, P.-M., Jacobeit, J., Fereday, D. R., Jones, P. D., Moberg, A., and Wanner, H. (2007). Long-term variability of daily North Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering. *Journal of Climate*, 20(16):4065–4095.

Pinto, J. G., Bellenbaum, N., Karremann, M. K., and Della-Marta, P. M. (2013). Serial clustering of extratropical cyclones over the North Atlantic and Europe under recent and future climate conditions. *Journal of Geophysical Research: Atmospheres*, 118(22).

Pinto, J. G., Ulbrich, S., Economou, T., Stephenson, D. B., Karremann, M. K., and Shaffrey, L. C. (2016). Robustness of serial clustering of extratropical cyclones to the choice of tracking method. *2016.*

Pinto, J. G., Ulbrich, U., Leckebusch, G., Spangehl, T., Reyers, M., and Zacharias, S. (2007). Changes in storm track and cyclone activity in three SRES ensem-

ble experiments with the ECHAM5/MPI-OM1 GCM. *Climate Dynamics*, 29(2-3):195–210.

Pinto, J. G., Zacharias, S., Fink, A. H., Leckebusch, G. C., and Ulbrich, U. (2009). Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO. *Climate dynamics*, 32(5):711–737.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G., Peubey, C., and Thépaut, J.-N. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29(11):4083–4097.

Prahl, B. F., Rybski, D., Kropp, J. P., Burghoff, O., and Held, H. (2012). Applying stochastic small-scale damage functions to German winter storms. *Geophysical Research Letters*, 39(6).

Rangel-Buitrago, N. and Anfuso, G. (2013). Winter wave climate, storms and regional cycles: the SW Spanish Atlantic coast. *International Journal of Climatology*, 33(9):2142–2156.

Rappaport, E. N., Franklin, J. L., Avila, L. A., Baig, S. R., Beven, J. L., Blake, E. S., Burr, C. A., Jiing, J.-G., Juckins, C. A., Knabb, R. D., et al. (2009). Advances and challenges at the National Hurricane Center. *Weather and Forecasting*, 24(2):395–419.

Renggli, D. (2011). *Seasonal predictability of wintertime windstorm climate over the North Atlantic and Europe*. Thesis, Freie Universität Berlin.

Renggli, D., Leckebusch, G. C., Ulbrich, U., Gleixner, S. N., and Faust, E. (2011). The skill of seasonal ensemble prediction systems to forecast wintertime wind-

storm frequency over the North Atlantic and Europe. *Monthly Weather Review*, 139(9):3052–3068.

Rivière, G., Arbogast, P., Lapeyre, G., and Maynard, K. (2012). A potential vorticity perspective on the motion of a mid-latitude winter storm. *Geophysical Research Letters*, 39(12).

Rivière, G., Arbogast, P., Maynard, K., and Joly, A. (2010). The essential ingredients leading to the explosive growth stage of the European wind storm Lothar of Christmas 1999. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 136(648):638–652.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, 27(6):2185–2208.

Sahai, A., Abhilash, S., Chattopadhyay, R., Borah, N., Joseph, S., Sharmila, S., and Rajeevan, M. (2015). High-resolution operational monsoon forecasts: an objective assessment. *Climate Dynamics*, 44(11-12):3129–3140.

Saunders, M. A. and Qian, B. (2002). Seasonal predictability of the winter NAO from North Atlantic sea surface temperatures. *Geophysical Research Letters*, 29(22).

Scaife, A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R., Dunstone, N., Eade, R., Fereday, D., Folland, C., and Gordon, M. (2014). Skillful long-range prediction of European and North american winters. *Geophysical Research Letters*, 41(7):2514–2519.

Schneider, T. and Griffies, S. M. (1999). A conceptual framework for predictability studies. *Journal of climate*, 12(10):3133–3155.

Schwierz, C., Köllner-Heck, P., Mutter, E. Z., Bresch, D. N., Vidale, P.-L., Wild, M., and Schär, C. (2010). Modelling European winter wind storm losses in current and future climate. *Climatic change*, 101(3-4):485–514.

Seierstad, I., Stephenson, D., and Kvamstø, N. (2007). How useful are teleconnection patterns for explaining variability in extratropical storminess? *Tellus A*, 59(2):170–181.

Semmler, T., Jung, T., and Serrar, S. (2016). Fast atmospheric response to a sudden thinning of Arctic sea ice. *Climate Dynamics*, 46(3-4):1015–1025.

Shannon, C. E. (1948). A note on the concept of entropy. *Bell System Tech. J*, 27(3):379–423.

Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell system technical journal*, 30(1):50–64.

Shapiro, M., Wernli, H., Bao, J.-W., Methven, J., Zou, X., Doyle, J., Holt, T., Donall-Grell, E., and Neiman, P. (1999). A planetary-scale to mesoscale perspective of the life cycles of extratropical cyclones: The bridge between theory and observations. In *The life cycles of extratropical cyclones*, pages 139–185. Springer.

Simmons, A. J. and Hoskins, B. J. (1979). The downstream and upstream development of unstable baroclinic waves. *Journal of the Atmospheric Sciences*, 36(7):1239–1254.

Simpson, R. H. and Saffir, H. (1974). The hurricane disaster potential scale. *Weatherwise*, 27(8):169.

SwissRe (2016). Natural catastrophes and man-made disasters in 2015: Asia suffers substantial losses. Report, Swiss Re, Zurich.

Tang, Y., Lin, H., Derome, J., and Tippett, M. K. (2007). A predictability measure applied to seasonal predictions of the Arctic Oscillation. *Journal of Climate*, 20(18):4733–4750.

Tierney, G., Posselt, D. J., and Booth, J. F. (2018). An examination of extratropical cyclone response to changes in baroclinicity and temperature in an idealized environment. *Climate Dynamics*, pages 1–18.

Torralba, V., Doblas-Reyes, F. J., MacLeod, D., Christel, I., and Davis, M. (2017). Seasonal climate prediction: A new source of information for the management of wind energy resources. *Journal of Applied Meteorology and Climatology*, 56(5):1231–1247.

Tripoli, G., Medaglia, C., Dietrich, S., and Mugnai, A. (2005). The 9-10 November 2001 Algerian flood. *Bulletin of the American Meteorological Society*, 86(9):1229.

Tsai, H.-C., Lu, K.-C., Elsberry, R. L., Lu, M.-M., and Sui, C.-H. (2011). Tropical cyclone–like vortices detection in the NCEP 16-day ensemble system over the western North pacific in 2008: Application and forecast evaluation. *Weather and Forecasting*, 26(1):77–93.

Ulbrich, U. and Christoph, M. (1999). A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing. *Climate dynamics*, 15(7):551–559.

Ulbrich, U., Fink, A., Klawa, M., and Pinto, J. (2001). Three extreme storms over Europe in December 1999. *Weather*, 56(3):70–80.

Vitolo, R., Stephenson, D. B., Cook, I. M., and Mitchell-Wallace, K. (2009). Serial clustering of intense European storms. *Meteorologische Zeitschrift*, 18(4):411–424.

Vrac, M. and Naveau, P. (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water resources research*, 43(7).

Walford, J. (2013). The Crimean War and fashion. `http://kickshawproductions.com/blog/?p=6560`. Accessed: 04/10/2018.

Walker, G. (1928). World weather. *Quarterly Journal of the Royal Meteorological Society*, 54(226):79–87.

Walker, M. (2011). *History of the meteorological office*. Cambridge University Press.

Walz, M. A., Befort, D. J., Kirchner-Bossi, N. O., Ulbrich, U., and Leckebusch, G. C. (2018a). Modelling serial clustering and inter-annual variability of European winter windstorms based on large-scale drivers. *International Journal of Climatology*.

Walz, M. A., Donat, M. G., and Leckebusch, G. C. (2018b). Large scale drivers and seasonal predictability of extreme wind speeds over the North Atlantic and Europe. *Journal of Geophysical Research: Atmospheres*.

Walz, M. A., Kruschke, T., Rust, H. W., Ulbrich, U., and Leckebusch, G. C. (2017). Quantifying the extremity of windstorms for regions featuring infrequent events. *Atmospheric Science Letters*, 18(7):315–322.

Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A., and Palmer, T. (2017). Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their

potential value for extreme event attribution. *Quarterly Journal of the Royal Meteorological Society*, 143(703):917–926.

Welker, C. and Martius, O. (2014). Decadal-scale variability in hazardous winds in northern Switzerland since end of the 19th century. *Atmospheric science letters*, 15(2):86–91.

Wheeler, D. (2001). The weather of the european atlantic seaboard during october 1805: An exercise in historical climatology. *Climatic change*, 48(2-3):361–385.

Wild, S., Befort, D. J., and Leckebusch, G. C. (2015). Was the extreme storm season in winter 2013/14 over the North Atlantic and the United Kingdom triggered by changes in the West Pacific Warm Pool? *Bulletin of the American Meteorological Society*, 96(12):S29–S34.

WMO (1970). The beaufort scale of wind force (technical and operational aspects). Bibliography: p.21-22.

Woollings, T. and Blackburn, M. (2012). The North Atlantic jet stream under climate change and its relation to the NAO and EA patterns. *Journal of Climate*, 25(3):886–902.

Woollings, T., Franzke, C., Hodson, D., Dong, B., Barnes, E. A., Raible, C., and Pinto, J. (2015). Contrasting interannual and multidecadal NAO variability. *Climate dynamics*, 45(1-2):539–556.

Zhang, W., Vecchi, G. A., Villarini, G., Murakami, H., Gudgel, R., and Yang, X. (2017). Statistical–dynamical seasonal forecast of western North pacific and East asia landfalling tropical cyclones using the gfdl flor coupled climate model. *Journal of Climate*, 30(6):2209–2232.

Appendix A

Appendix

A.1 Footprint of windstorm Vivian

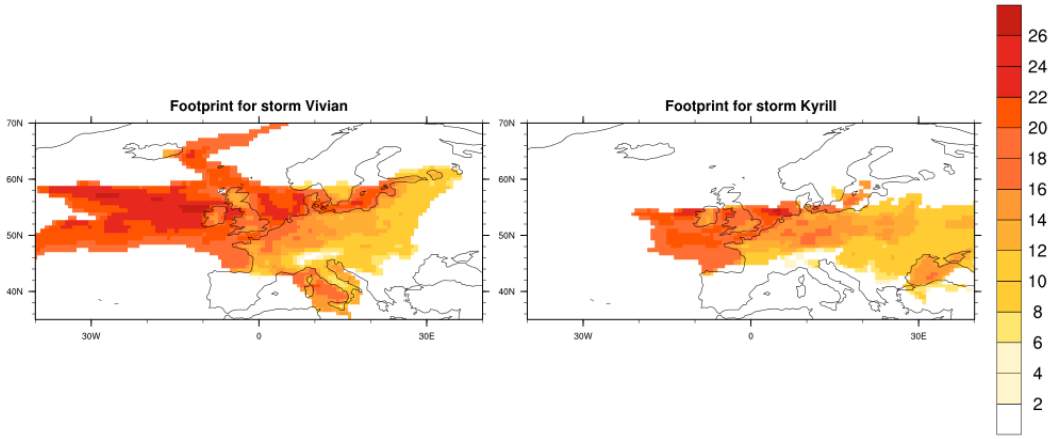


Figure A.1.1: *Footprint of maximum wind speed for storm Vivian and Kyrill. The very extreme wind speeds over the Central Atlantic Ocean for Vivian are responsible for the very large DI-SSI value compared to all the other storms which are compared in Table 5.4.2. This is due to the fact that the DI-SSI shows the same magnitude for wind speeds of the same extremeness.*

A.2 10 leading rotated EOFs of ERA–20C

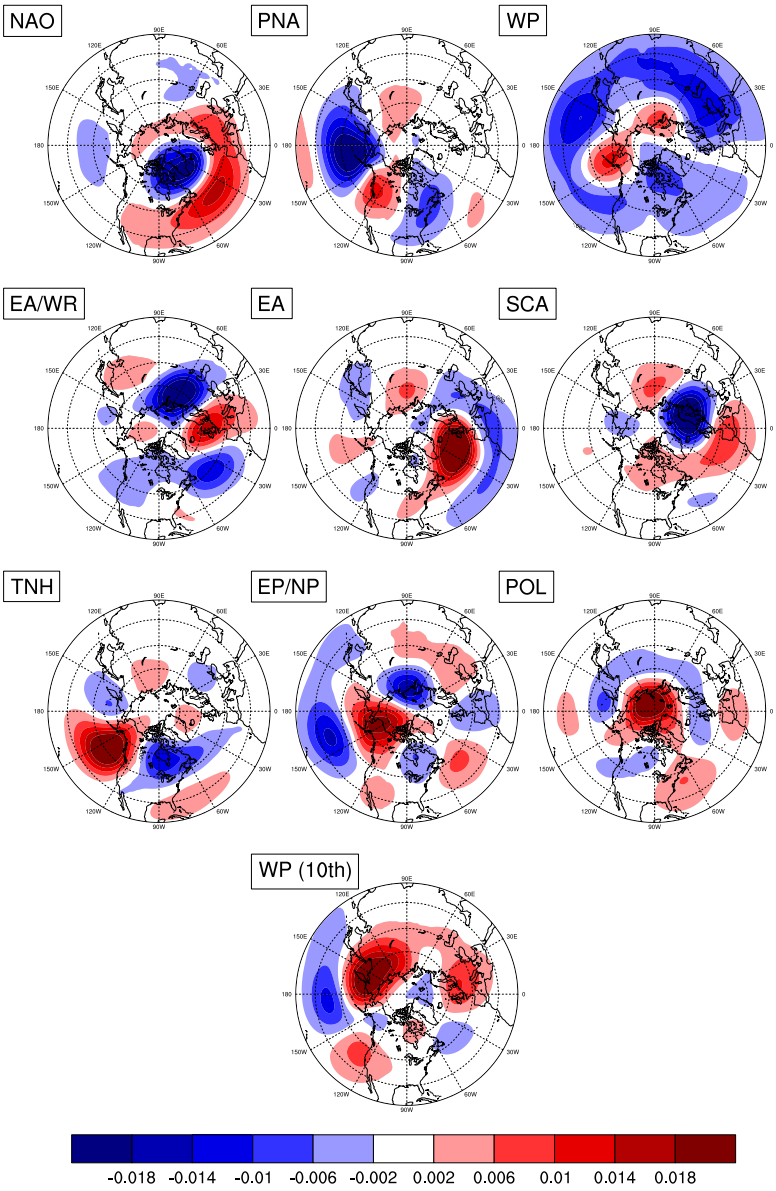


Figure A.2.1: *Leading 10 rotated EOFs as calculated from ERA–20C and used as predictors for the statistical model for modelling serial clustering of windstorms*

A.3 Correlation of extreme wind speeds for the non-aggregated time series

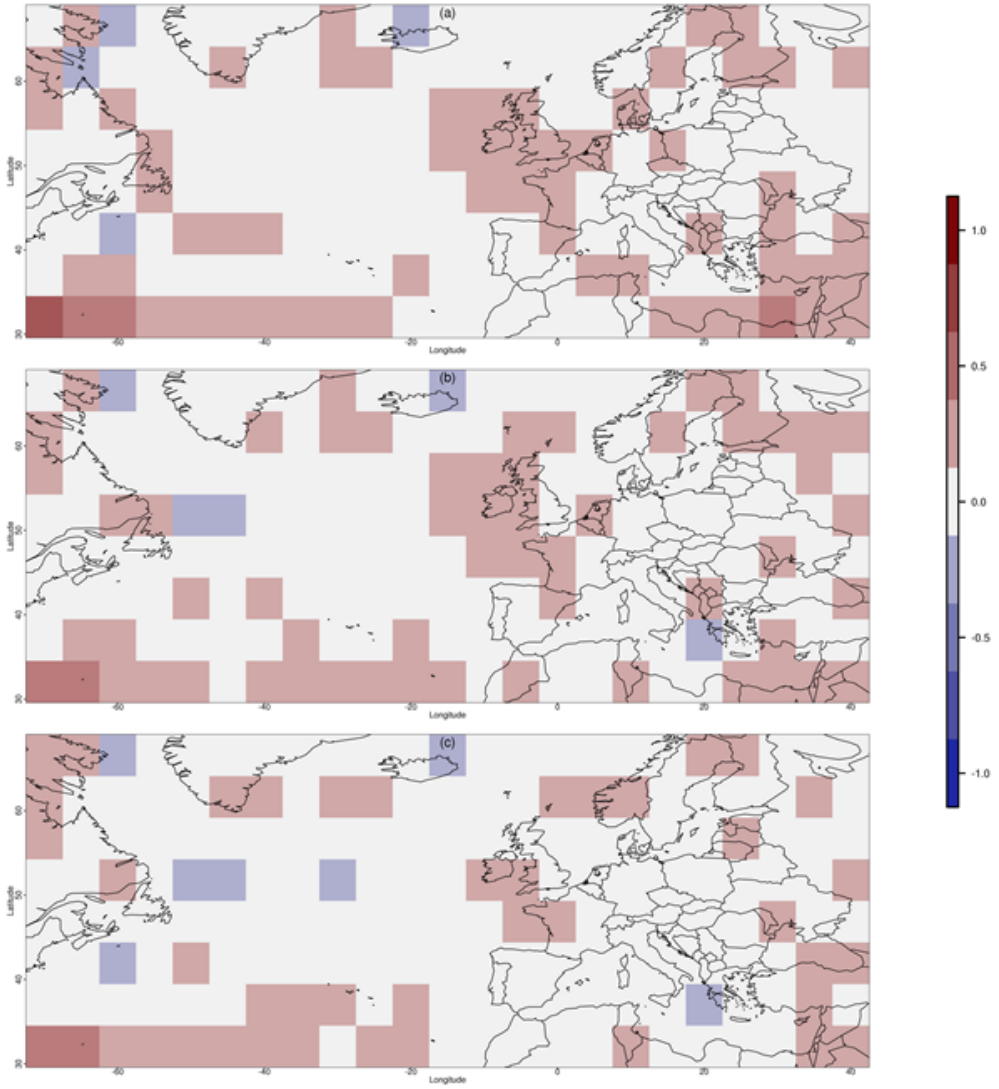


Figure A.3.1: *Correlation maps for time series of the 90^{th} (a), 95^{th} (b) and 98^{th} (c) percentiles of wind speeds without using the aggregated variable between System 4 and ERA-Interim. None of the correlations are significant at the 5% confidence level.*

A.4 Large-scale drivers selected based on the System 4 internal EOF time series

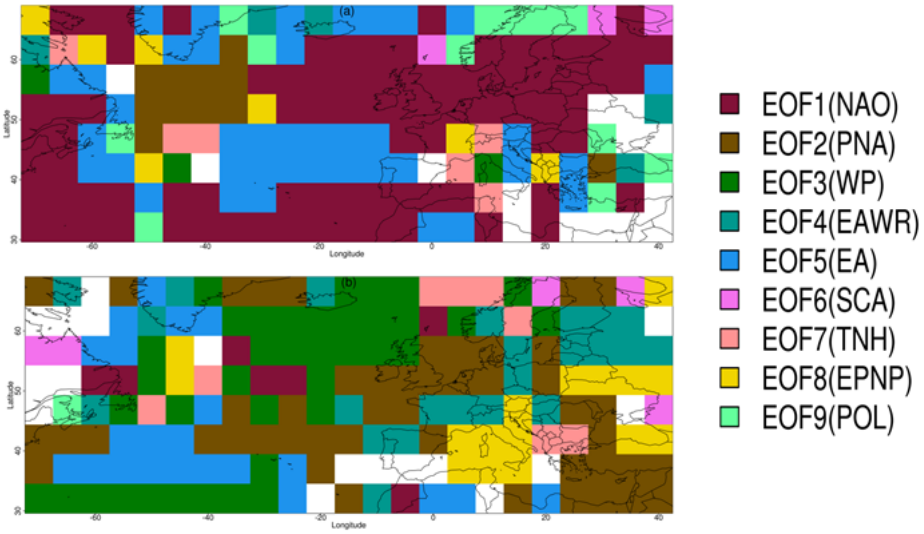


Figure A.4.1: *MSLP variability drivers associated with the largest absolute regression coefficient for every grid cell for explaining the inter-annual variability of $I_{k,y,95}$ when using the System 4 internal EOF eigenvectors to compute the time series. Equivalent figure to Figure 4.4.4. If no significant driver could be determined the grid cell remains white. Winning drivers for ERA-Interim are shown in the top figure (a), winning drivers for System 4 are shown in the bottom figure (b).*

A.5 Description of predictors for the statistical model

Table A.5.1: *Overview of the potential large scale drivers considered for the development of the statistical model of winter wind storms over Europe*

Index Name	Index Source	Additional Information
QBO 30hPa	Observations	Broennimann et al. (2007)
QBO 70hPa	Observations	Broennimann et al. (2007)
AMO	ERA-20C (SSTs)	SSTs over North Atlantic (85W-5W 10-70N) for DJF season
HSI (Horse-Shoe-Index)	ERA-20C (SST)	Index calculation (DJF) according to Renggli, 2011
SSTS	ERA-20C (SST)	Southern box of HSI
Tdif.Nam	ERA-20C (SST)	Index calculation according to Wild et al. (2015)
W.Atl T	ERA-20C (SST)	West Atlantic box from Tdif.Nam (DJF)
NINO3.4	ERA-20C (SST)	Region 170-120W, 5S-5N (DJF)
NAO.Is.Li	ERA-20C (MSLP)	Season (DJF) Station based NAO index
PDO	ERA-20C (SST)	1st EOF of the DJF Pacific SSTs anomalies (120E-105W, 20-65N).
West Pacific	ERA-20C (GPH 700hPa)	3rd rotated EOF (rEOF) from 700hPa GPH for northern hemisphere.
PNA	ERA-20C (GPH 700hPa)	2nd rEOF from 700hPa GPH for northern hemisphere.
West.Pacific.10th EOF	ERA-20C (GPH 700hPa)	10th rEOF from 700hPa GPH for northern hemisphere.
EA.WR	ERA-20C (GPH 700hPa)	4th rEOF from 700hPa GPH for northern hemisphere.
EA	ERA-20C (GPH 700hPa)	5th rEOF from 700hPa GPH for northern hemisphere.
SCA	ERA-20C (GPH 700hPa)	6th rEOF from 700hPa GPH for northern hemisphere.
TNH	ERA-20C (GPH 700hPa)	7th rEOF from 700hPa GPH for northern hemisphere.
EP.NP	ERA-20C (GPH 700hPa)	8th rEOF from 700hPa GPH for northern hemisphere.
POL	ERA-20C (GPH 700hPa)	9th rEOF from 700hPa GPH for northern hemisphere.
Sea Ice	ERA-20C (sea ice)	Sea ice cover 40-90N for DJF